

**TRANSCRIPTIONAL REGULATION DURING THE PAPILLOMAVIRUS LIFE
CYCLE AND ELIMINATION OF INFECTION USING HOMOLOGOUS
RECOMBINATION**

by

Andrew Robert Carson

BS Microbiology, Molecular Biology and Biochemistry, University of Idaho, 1999

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

University of Pittsburgh

2006

University of Pittsburgh

School of Medicine

This dissertation was presented

by

Andrew Robert Carson

It was defended on

March 23, 2006

And approved by

Dr. Saleem Khan, Professor, Molecular Genetics and Biochemistry

Dr. Neal DeLuca, Professor, Molecular Genetics and Biochemistry

Dr. Graham Hatfull, Professor, Biological Sciences

Dr. Christine A. Milcarek , Professor, Immunology

Dr. Xiao Xiao, Associate Professor, Orthopedic Surgery

TRANSCRIPTIONAL REGULATION DURING THE PAPILLOMAVIRUS LIFE CYCLE AND ELIMINATION OF INFECTION USING HOMOLOGOUS RECOMBINATION

Andrew Robert Carson, PhD

University of Pittsburgh, 2006

Human Papillomaviruses (HPVs) require terminal differentiation of the host cell to produce infectious virions. The process of viral maturation involves a variety of changes in the expression/activity of host proteins that lead to high-level replication of the viral genome and expression of the late viral genes. Although the late promoter regions of HPV-16 are still not fully characterized, differentiation-dependent regulation of viral genes is thought to involve changes in the binding of host cell transcription factors to the viral promoter and regulatory regions. Currently, little is known about specific cellular transcription factors involved in this process. We have used the Panomics TransSignal Protein/DNA array to identify changes in the levels of cellular transcription factors during methylcellulose-induced differentiation of W12 (20863) cells containing HPV-16. We then identified the differentially expressed transcription factors that specifically bind to HPV-16 promoters. We have validated the results obtained from the Panomics array by Western blot analysis and with chromatin immunoprecipitation. This approach identified approximately thirty transcription factors, many of which represent novel viral DNA-host protein interactions. At present, no treatments exist that effectively target and eliminate papillomaviruses (PVs) from infected cells or prevent its replication. We are employing a strategy to prevent virus replication in PV-infected cells through the conditional expression of the herpes simplex virus type 1 thymidine kinase (TK) gene. Expression of TK in this system is expected to be triggered by a homologous recombination event between the endogenous PV genome and a nonexpressing TK gene cassette, which is expected to change the nonexpressing cassette into a form that expresses TK. Various constructs were generated to express the TK in the above manner. Transfection of cell lines with a TK nonexpressing plasmid did not result in TK production. However, cotransfection of cell lines with PV plasmids along with the above TK construct containing PV sequences resulted in TK expression as shown by Northern and Western blot analyses. We also developed a TK expression cassette utilizing an adeno-associated virus (AAV) vector. Delivery of the cassette by AAV to PV-infected cells resulted in TK expression, and ganciclovir treatment resulted in efficient killing of these cells.

TABLE OF CONTENTS

<u>1.0</u>	<u>GENERAL INTRODUCTION</u>	1
<u>1.1</u>	<u>PAPILLOMAVIRUS LIFE CYCLE</u>	2
<u>1.2</u>	<u>HUMAN PAPILLOMAVIRUSES AND CANCER</u>	6
<u>1.3</u>	<u>PAPILLOMAVIRUS DNA AS A TARGET FOR RECOMBINATION</u>	9
<u>1.4</u>	<u>ALTERNATIVE POLYADENYLATION SIGNAL USAGE</u>	12
<u>1.5</u>	<u>HERPES SIMPLEX VIRUS TYPE 1 THYMIDINE KINASE</u>	13
<u>1.6</u>	<u>ADENO-ASSOCIATED VIRUS AS A VECTOR</u>	14
<u>1.7</u>	<u>HPV-16 TRANSCRIPTION</u>	16
<u>1.8</u>	<u>THESIS OUTLINE</u>	22
<u>2.0</u>	CHARACTERIZATION OF TRANSCRIPTION FACTOR BINDING TO HUMAN PAPILLOMAVIRUS TYPE 16 DNA DURING CELLULAR DIFFERENTIATION 25	
<u>2.1</u>	<u>INTRODUCTION</u>	26
<u>2.2</u>	<u>MATERIALS AND METHODS</u>	30
<u>2.3</u>	<u>RESULTS</u>	37
<u>2.4</u>	<u>DISCUSSION</u>	48
<u>3.0</u>	<u>A DNA RECOMBINATION-BASED APPROACH TO ELIMINATE PAPILLOMAVIRUS INFECTION</u>	52
<u>3.1</u>	<u>INTRODUCTION</u>	53
<u>3.2</u>	<u>MATERIALS AND METHODS</u>	56
<u>3.3</u>	<u>RESULTS</u>	61
<u>3.4</u>	<u>DISCUSSION</u>	74
<u>4.0</u>	<u>GENERAL DISCUSSION</u>	79
<u>4.1</u>	<u>CHARACTERIZATION OF TRANSCRIPTION FACTORS BINDING TO HPV-16 DURING CELLULAR DIFFERENTIATION</u>	80
<u>4.2</u>	<u>FUTURE DIRECTIONS IN THE STUDY OF HPV-16 TRANSCRIPTION</u>	84
<u>4.3</u>	<u>OUTLOOK ON A RECOMBINATION-BASED APPROACH TO ELIMINATE PAPILLOMAVIRUS INFECTION PAPILLOMAVIRUS LIFE CYCLE</u>	87
<u>4.4</u>	<u>FUTURE OF RECOMBINATION-BASED APPROACHES TO ELIMINATE VIRAL INFECTIONS</u>	98
<u>5.0</u>	<u>BIBLIOGRAPHY</u>	100

LIST OF TABLES

Table 1. Changes in transcription factor binding to HPV-16 promoters during cellular differentiation	40
---	-----------

LIST OF FIGURES

<u>Figure 1. The human papillomavirus life cycle</u>	3
<u>Figure 2. The organization of circular HPV DNA and its integration into host-cell DNA</u>	7
<u>Figure 3. HPV-16 transcription</u>	17
<u>Figure 4. Late promoter(s) of HPV-16</u>	20
<u>Figure 5. 20863 cells express the differentiation marker keratin K10 upon treatment with methylcellulose and C8</u>	38
<u>Figure 6. Putative TF binding sites in the HPV-16 promoters and upstream regions</u>	43
<u>Figure 7. Western blot analysis of TFs in differentiated and undifferentiated cells</u>	44
<u>Figure 8. Chromatin immunoprecipitation (ChIP) and quantitative PCR analysis of TF binding</u>	46
<u>Figure 9. Diagram of various TK constructs used in this study</u>	62
<u>Figure 10. Western blot analysis of TK expression</u>	64
<u>Figure 11. Northern blot analysis of COS-1 cells transfected with the various TK plasmids</u>	65
<u>Figure 12. Recombination-dependent expression of FLAG-TK in cells transfected with various constructs</u>	69
<u>Figure 13. Recombination-dependent TK expression in BPV-positive cells upon infection with an AAV-TK vector</u>	71
<u>Figure 14. GCV inhibits the growth of BPV-positive cells infected with an AAV-TK vector</u>	73
<u>Figure 15. Model of homologous recombination leading to TK expression</u>	88

Chapter 1

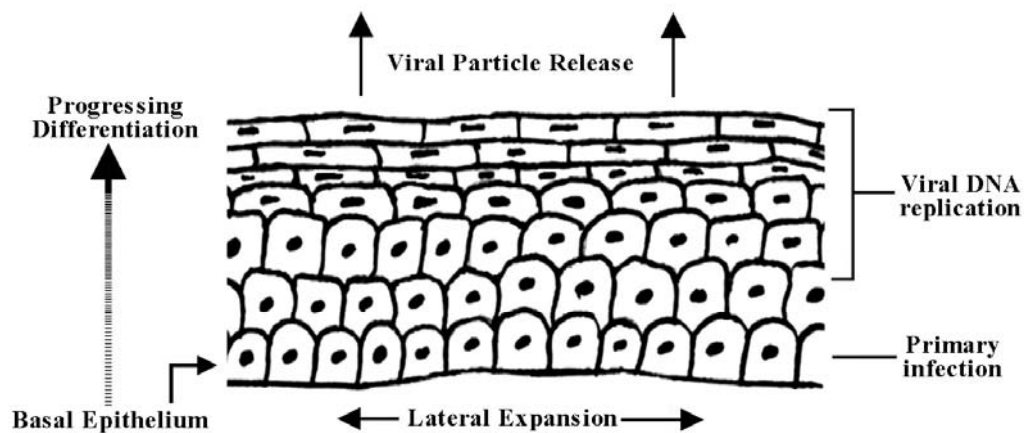
GENERAL INTRODUCTION

1.1 PAPILLOMAVIRUS LIFE CYCLE

Papillomaviruses (PVs) are small, nonenveloped, double-stranded DNA tumor viruses with genomes of about 8,000 base pairs (zur Hausen, 2002; Hebner & Laimins, 2005). PVs are a large and diverse group of viruses, with each virus displaying a tropism for certain tissues in a specific vertebrate host. Some animal PVs, such as bovine papillomavirus type 1 (BPV-1), are well characterized and serve as important models of papillomavirus biology (zur Hausen, 2002; Hebner & Laimins, 2005). Nearly 200 types of human papillomaviruses (HPVs) have been identified to date and are grouped according to their tropism towards specific epithelial cells. All Papillomaviruses infect epithelial cells and all require terminal differentiation of the host cell to produce infectious virions (zur Hausen, 2002; Hebner & Laimins, 2005).

For productive infection to occur, PVs must gain access to the basal epithelium through an abrasion or other type of cut in the skin. For example, HPV types 1 and 2 must gain access to cutaneous basal epithelium, while types 6 and 16 must gain access to mucosal basal epithelium (zur Hausen, 2002; Hebner & Laimins, 2005). The virus then enters the host cell through a mechanism that is not entirely understood, but probably involves cellular integrins (zur Hausen, 2002; Hebner & Laimins, 2005). Following viral uncoating and transport to the nucleus, an initial burst of replication brings the viral genome to a copy number of about 50 per cell (this copy number is type dependent) (zur Hausen, 2002; Hebner & Laimins, 2005). The viral genomes are maintained episomally at these low copy numbers for long periods of time in the basal epithelium, where the virus replicates in synchrony with the S-phase of the host cell (zur Hausen, 2002; Hebner & Laimins, 2005). This life cycle is illustrated in Fig. 1.

Fig 1. The human papillomavirus life cycle



Infection occurs in the basal epithelium. The infected cell divides and spreads laterally. Progeny also migrate into the differentiating suprabasal layer, viral genes are activated, viral DNA is replicated, and capsid proteins assemble to produce infectious virions. Virus particles are released at the surface.

The viral proteins E1 and E2 are expressed at this time and have a number of important functions. E1 is a highly conserved protein across all HPVs, and has been shown to bind to the viral origin of replication, and act as a helicase to unwind the DNA in the region (zur Hausen, 2002; Hebner & Laimins, 2005; Wilson *et al*, 2002). E1 is absolutely required for the initiation of replication in all PV types. The origins of replication of PVs contain an E1 binding site that consists of an 18-bp palindromic sequence (Holt & Wilson, 1995; Holt *et al*, 1994; Sun *et al*, 1996). E2 also stimulates replication in all PV types and is required for the replication of some HPV types (such HPV-16) but not required in other types (such as HPV-1) (Gopalakrishnan & Khan, 1994). There are multiple E2 binding sites in the origin of replication of most HPVs and it has been shown that E1 and E2 bind cooperatively to this origin in some HPV types (Van Horn *et al*, 2001; Chiang *et al*, 1992). E2 binding sites are also located in viral promoters, where the protein acts to regulate transcription (zur Hausen, 2002; Hebner & Laimins, 2005). E2 has been shown to suppress transcription of the viral oncogenes E6 and E7, which is of particular importance in HPV related cancers (discussed later).

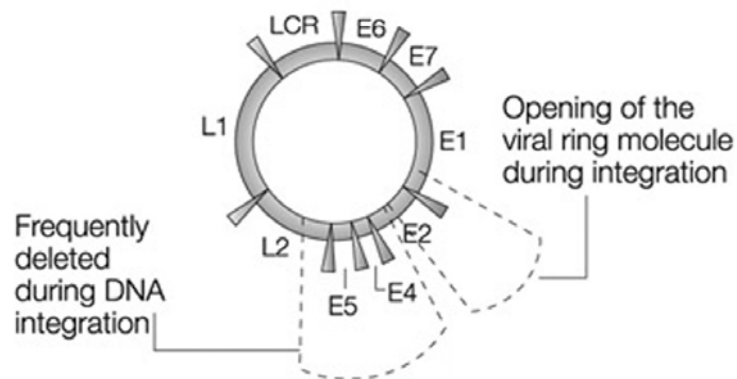
Upon the differentiation of the host cell into the spinous layers of the epithelium, there is increased expression of E1 and E2 and the virus replicates to high levels (zur Hausen, 2002; Hebner & Laimins, 2005). Transcription of the potent oncoproteins E6 and E7 is also increased at this stage of differentiation. E6 interacts with the p53 tumor suppressor and targets it for degradation via the proteasome pathway (zur Hausen, 2002; Hebner & Laimins, 2005). E6 also interacts with many other host cell proteins including the tumor suppressors hDlg and hScrib (Kiyono *et al*, 1997; Lee, Weiss & Javier, 1997; Nakagawa & Huibregtse, 2000), the replication protein Mcm7 (Kuhne & Banks, 1998;

Kukimoto *et al*, 1998), transcriptional factors c-myc and CBP/p300 (Gross-Mesilaty *et al*, 1998; Patel *et al*, 1999; Zimmermann *et al*, 1999). E7 binds the RB tumor suppressor, preventing the binding of RB to E2F and releasing E2F to promote the progression of the cell cycle (zur Hausen, 2002; Hebner & Laimins, 2005). Normally, the binding and sequestration of E2F by the hypophosphorylated form of RB acts as an important checkpoint in the cell cycle. E7 has also been shown to associate with the cell-cycle control protein p21 and histone deacetylase HDAC (Finzer *et al*, 2001). E5 is expressed in basal and spinous epithelium and may play a role in transformation, although not as prominent a role as the E6 or E7 proteins. It has been shown that E5 can activate epidermal growth factor receptor (EGFR) through binding to the 16kD subunit, an ATPase protein pump, and initiates biochemical cascades that lead to overexpression of a variety of protooncogenes. Additionally, E5 can inhibit the expression of tumor suppressor gene p21 (Tsai & Chen, 2003). E4 expression is also induced upon differentiation into the spinous layer where it has been shown to alter mitotic signals and destabilizes keratin (zur Hausen, 2002; Hebner & Laimins, 2005). Importantly, the major and minor capsid proteins L1 and L2 are also produced in the spinous layer. Mature virions are assembled in the granular layers of epithelium and shed with the sloughing of the stratum corneum (zur Hausen, 2002; Hebner & Laimins, 2005).

1.2 HUMAN PAPILLOMAVIRUSES AND CANCER

Human Papillomaviruses are associated with over 99% of cervical cancers worldwide and have also been associated to a lesser degree with other cancers, such as oral cancer (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). Cervical cancers are strongly linked to certain HPV types, such as types 16, 18, 31, and 45, which are referred to as high risk HPV types (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). Low-risk types, such as types 1, 6, and 11, also infect cervical epithelium but are associated with low-grade cervical lesions, and almost never with cancer (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). The HPV genome often integrates into the host chromosome near the viral E2 locus (see Fig 2) in malignancies associated with HPV (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). The mechanism for this integration is unknown, and it is unclear if the integration event is targeted. The site of HPV disruption in these cancers is almost always in the E2 ORF, but the exact location within that region, as well as in the host genome, is seemingly random (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). Integration in the E2 ORF prevents the expression of E2, and the loss of E2 has been shown to result in transcriptional upregulation of the potent oncogenes E6 and E7 (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). One hypothesis is that the viral genome integrates randomly, but there is a growth advantage in cells which have no E2 expression and overexpression oncogenes as a result.

Fig 2. The organization of circular HPV DNA and its integration into host-cell DNA



The human papillomavirus (HPV) genome contains eight open reading frames (E6, E7, E1, E2, E4, E5, L2 and L1). In the course of cancer development, the viral molecule often integrates into host-cell DNA. The ring molecule is most often opened within the E2 open reading frame, disrupting that gene. Part of E2 as well as adjacent open reading frames (E4, E5, and L2) are often deleted after integration.

It should be noted that viral integration prevents productive replication, and represents a negative selection for viruses that cause cancer. This is demonstrated by the fact that a majority of people who are persistently infected with HPV do not develop HPV-related cancer in their lifetime (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). This is consistent with the hypothesis that integration is a low probability “accident” which takes years to occur and that the oncogenic potential of HPV is greatly increased by its ability to cause persistent infections. It is known that malignancies associated with HPV infection often occur several years after the primary infection (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). It is also known that HPV lesions may persist for decades without clearance or regression, and surgical treatment does not prevent recurrence (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). At present, no effective treatment is available to target and eliminate HPV from persistently infected cells.

1.3 PAPILLOMAVIRUS DNA AS A TARGET FOR RECOMBINATION

During persistent infection, large populations of host cells harbor episomal HPV DNA at various copy numbers. This property can be utilized as a molecular target through the use of homologous recombination to trigger the expression of the suicide gene Herpes Simplex Virus type 1 Thymidine Kinase (TK). Homologous recombination is just one mechanism by which cells repair double-stranded DNA breaks (DSBs), which can be caused by exposure to ionizing radiation, DNA-damaging agents, or stalled replication forks (Haber, 1999). Because DSBs in chromosomes are lethal, cells have developed several pathways that compete for repair of these double-stranded breaks, one of which is homologous recombination. Homologous recombination can occur through a process known as gene conversion (Haber, 1999). Classically, this process begins with the resection of the 5' end of the DNA at the DSB. The resulting 3' overhangs can then invade the donor strand through extensive base-pairing to complementary sequences in the donor strand (Haber, 1999). This strand invasion is catalyzed by the RAD51 protein. RAD52 strongly stimulates strand invasion, possibly by binding to DSBs and recruiting RAD51 to these sites (McIlwraith *et al*, 2000). The single-stranded binding protein RPA also strongly stimulates single-strand annealing by RAD51, possibly by removing secondary structure in the 3' overhang which might repress RAD51 binding (McIlwraith *et al*, 2000). Following strand invasion, DNA polymerization occurs at the 3' end of the invading DNA strand using the donor strand as a template. The newly synthesized strand of DNA then anneals to homologous sequences on the other end of the DSB to accurately repair the junction at the site of the break (Haber, 2000). This process of strand exchange

forms a Holliday structure that may result in crossing over when the structure is resolved (Haber, 2000).

Single-stranded annealing, like gene conversion, begins with the resection of the 5' ends surrounding the break. Unlike gene conversion, resection of the 5' ends at the double-stranded break exposes small homologous sequences originally flanking the break. These sequences then anneal as single-stranded DNA and are ligated together to create a deletion at the site of the double-stranded break (Haber, 2000). Non-homologous end joining also competes with homologous recombination for repair of double-stranded breaks. In this process, free ends resulting from a double-stranded break are joined through one of multiple mechanisms, none of which involves extensive homology between the sequences. The result is that the break is repaired, but the junction is often inaccurately reconstituted and frequently contains deletions, misalignments, and additions of small sequences (Haber, 2000).

It should be noted that homologous recombination occurs frequently between episomal sequences. This was demonstrated in numerous experiments in which plasmids containing homologous sequences were cotransfected into mammalian cells and recombination between the plasmids was detected using antibiotic resistance, southern blot, and restriction enzyme analysis. Intramolecular and intermolecular homologous recombination between identical sequences co-transfected into immortalized cells occurs in roughly 0.2% to 1% of recovered plasmids when the recombination donor and target sequences are co-transfected at nearly equimolar ratios (Nickoloff & Reynolds, 1990; Thyagarajan *et al*, 1996; Rouet *et al*, 1994; Seidman, 1987). Consistent with the idea that DSBs induce homologous recombination; the frequency of recombination can be

dramatically increased by linearizing one or both of the plasmids undergoing recombination near the sites of recombination prior to transfection. Cutting one plasmid increases the recombination frequency 2-7 fold (Ayers *et al*, 1985) and cutting both fragments near the recombination sites results in frequencies of 10% to 30% (Ayers *et al*, 1985; Rouet *et al*, 1994). Homologous recombination between replicating plasmids occurs continuously over a 48-hour period (Ayers *et al*, 1985), which argues that increased recombination frequencies could be seen over extended periods of time. If self-replicating plasmids are allowed to recombine for a one-month period, over 50% of the plasmids undergo a type of irreversible intramolecular homologous recombination (Kitamura *et al*, 1990). These frequencies are in contrast with chromosomal homologous recombination, which occurs with frequencies of less than 1 in 10 million cells in the absence of a DSB and in only 0.2-1% of cells where a break has been induced at the site of recombination in the chromosome (Taghian & Nickoloff, 1997). This is of particular interest because, as mentioned previously, papillomavirus replicates as an extrachromosomal element in the host cell.

1.4 ALTERNATIVE POLYADENYLATION SIGNAL USAGE

In order to utilize recombination to trigger the expression of the suicide gene, TK, a very tightly regulated expression system must be developed. We have chosen to regulate the expression of TK through alternative splicing and polyadenylation signal usage. The expression of many natural genes is known to be regulated by alternative poly(A) site selection and this selection is known to be effected by both the position and relative strength of the poly(A) sites (reviewed in Edwards-Gilbert *et al*, 1997). Among identical poly(A) signals, the 5' site is favored over a more distal downstream site (Demone & Cole, 1988). Also, when a strong poly(A) site precedes a weak one by a short distance, less than 5% of the processed mRNAs terminate at the weak site (Batt *et al*, 1994). It is also known that poly(A) site usage can be affected by surrounding splice sites. Cooke *et al* (1999) have shown that a 5' splice site, if not followed by a 3' splice site, can interfere with downstream polyadenylation. Poly(A) site usage in alternatively spliced regions can also be suppressed through exon skipping of the poly(A) signal. Exon skipping has been shown to decrease when the 5' splice site following the skipped exon is strengthened, but increases when the 5' splice site is weakened or eliminated (Kuo *et al*, 1991).

1.5 HERPES SIMPLEX VIRUS TYPE 1 THYMIDINE KINASE

HSV-1 Thymidine Kinase (TK), in contrast to mammalian thymidine kinases, converts the nontoxic nucleoside analogue ganciclovir to a monophosphorylated nucleotide analogue. In the cell, the monophosphorylated form of ganciclovir is further phosphorylated to the triphosphate form. This form of ganciclovir is incorporated into the growing strand of DNA during replication by cellular polymerases. Incorporation causes premature chain termination, preventing replication of the host genome and results in the loss of viability of the host cell (Salomon *et al*, 1995). HSV-1 TK is, therefore, an effective suicide gene when used in conjunction with ganciclovir and has been used successfully in many gene therapy approaches. Notably, Caruso *et al* (1995) have used TK under the transcriptional control of a truncated form of the HIV-1 LTR to induce TK expression in transfected tissue culture cells expressing the LTR-specific viral elongation factor Tat. They found that the process of transfecting the TK expression construct and treating the transfected cells with acyclovir (a drug similar to ganciclovir) was effective at killing cells expressing Tat, with minimal effects on non-expressing cells (Caruso *et al*, 1995). This conditional expression of TK is useful in only fractions of HIV-1 infected cells, however, because both the expression of Tat and the transcriptional processivity mediated by Tat can vary greatly in infected cells *in vivo*, with some infected cells having no measurable Tat activity (Adams *et al*, 1999). Although the concept investigated by Salomon *et al*, Caruso *et al*, and others shows great efficacy, it is not based on a molecular target that must be present in all infected cells. Obviously, a viral genome must be present in infected cells during productive, latent, or persistent infections.

1.6 ADENO-ASSOCIATED VIRUS AS A VECTOR

AAV is a small (4.7 Kb) single-stranded DNA virus that is being increasingly used as a gene-delivery vector (Hirata & Russell, 2000; Sun *et al*, 2000; Xiao *et al*, 1998). AAV has many attractive features for use in delivering recombination-dependent TK expression cassettes. Of particular relevance is its ability to efficiently infect many types of dividing and non-dividing cells, the low immune response generated to it, and its persistent expression of transgenes (Sun *et al*, 2000). Importantly, AAV is highly recombinogenic both at the terminal repeats that flank the genome and throughout the genome (Hirata & Russell, 2000; Sun *et al*, 2000). Though no extrachromosomal gene replacement studies have been performed, chromosomal gene replacement studies using AAV by Hirata and Russell (2000) show a 3-4 log increase in recombination frequencies over those achieved by conventional methods. Wild-type AAV-2 has also been found in significant numbers of cervical epithelial cells from patients (Venturoli *et al*, 2001), indicating that AAV can infect the same types of cells that high risk HPV types infect. Raft culture studies of human epithelial cells have demonstrated that AAV and HPV can also co-infect the same cell (Meyers *et al*, 2001). This is evidence that recombinant AAV could effectively deliver a recombination-dependent TK expression cassette to a patient *in vivo*. As mentioned previously, recombinant AAV can transduce many types of human cells both *in vivo* and *in vitro*, but AAV does not efficiently transduce many transformed rodent cell lines (Hansen *et al*, 2001; Russel *et al*, 1995). Unfortunately, these same cell lines harbor the BPV episomality at high copy number. Pretreatment of these cells with the ribonucleotide reductase inhibitor hydroxyurea (HU) increases transduction efficiency dramatically (Hansen *et al*, 2001). Although the mechanism is

still unknown, this increase in transduction efficiency may be due to alteration of trafficking in these cell lines and not due to the classical HU activities such as S-phase arrest and/or induction of DNA repair functions (Hansen *et al*, 2001; Russel *et al*, 1995).

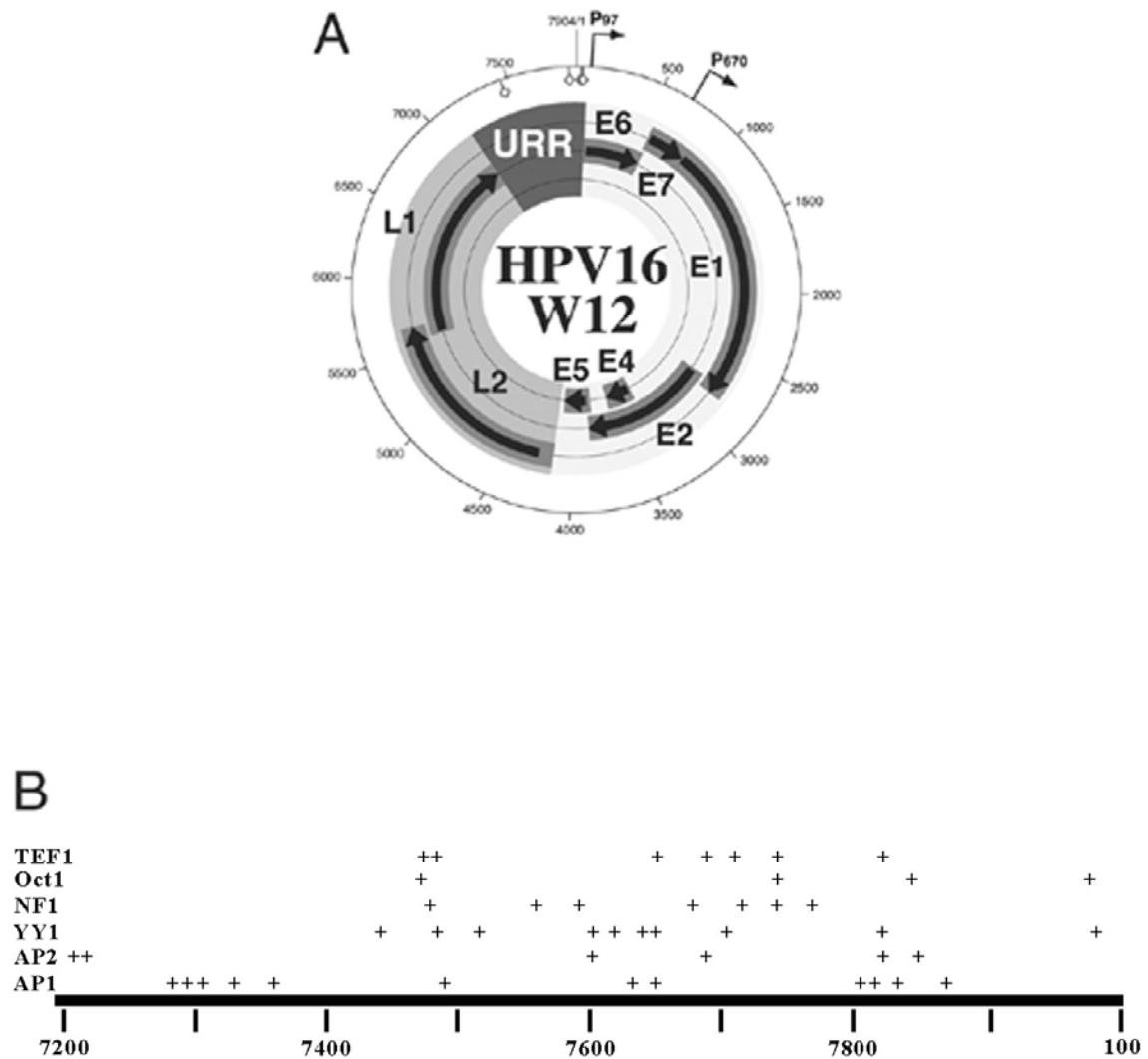
1.7 HPV-16 TRANSCRIPTION

As mentioned previously, Papillomavirus requires terminal differentiation of the host cell to produce infectious virions. The process of viral maturation involves a regulated and orderly increase in expression of at least six proteins (zur Hausen, 2002; Hebner & Laimins, 2005). Amazingly, HPV-16 only expresses one transcription factor, E2, and E2 activity is dispensable for the upregulation of gene products upon differentiation. It is likely that HPV has evolved to utilize naturally occurring changes in host cell transcription factor expression and activation to express viral genes in an ordered cascade. This stratagem enables HPV-16 to regulate the expression of at least eight products from a small genome comprised of about 90% open reading frames. To fit so many ORFs in such a small genome, HPV-16 makes use of a complex pattern of alternative splicing, alternative polyadenylation signal usage, alternative translational start sites, and multiple promoters, as illustrated in Fig. 3 (zur Hausen, 2002; Hebner & Laimins, 2005).

Early p97 promoter

The most active and best characterized promoter of HPV-16 is the p97 promoter (major transcriptional start site at base pair 97). This promoter is responsible for the expression of E6, E7 and probably responsible for E1, E2, E4 and E5 as well (zur Hausen, 2002; Hebner & Laimins, 2005). There is uncertainty about many of these genes because the 5' end of many mRNA species have not been fully mapped, and complex splicing patterns as well as alternative translational start sites make it difficult to determine which transcripts express specific proteins. Through alternative splicing, the

Fig 3. HPV-16 transcription



(A) Circular map depicting the early and late genes of the viral genome.
 (B) Linear map of the entire upstream regulatory regions. Potential binding sites for cellular transcription factors are indicated by (+) symbols.

p97 promoter may also express L1 and L2. It has been found that p97 activity increases 1.5 to 3 fold upon differentiation of the host cell (Apt *et al*, 1996; Grassmann *et al*, 1996; Yukawa *et al*, 1996) and correlates to increases in expression of early genes. As mentioned previously, E6 and E7 expression is dramatically induced upon differentiation (zur Hausen, 2002; Hebner & Laimins, 2005). It remains unclear how an increase in E2 can accompany an increase in E6 and E7, which has been shown to be suppressed by E2 expression from the p97 promoter in undifferentiated cells. Also, L1 and L2 expression is undetectable before differentiation, while easily detectable afterwards (zur Hausen, 2002; Hebner & Laimins, 2005). This is likely more than a 1.5-3 fold increase in transcription. While this could be a result of altered splicing patterns or polyadenylation signal usage, it argues that other promoters may be utilized to express these late genes.

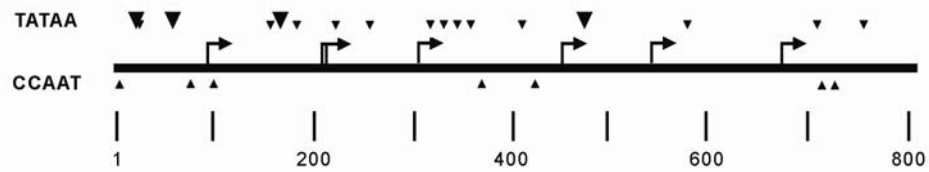
It is clear that understanding how p97 promoter activity changes during differentiation is crucial to understanding the life cycle of the virus. Many common cellular transcription factors such as AP1 (Apt *et al*, 1993; Chan *et al*, 1990), NF1 (Ai *et al*, 1999), Oct-1 (Chan *et al*, 1990), SP1 (Flores *et al*, 1997), and YY1 (Ciccolini *et al*, 1994) have been shown to activate or inhibit the p97 promoter, as shown in Fig. 3. The expression or activity of many of these common transcription factors are not known to change dramatically during differentiation. A few differentiation specific transcription factors, such as CDP (Lee & Haber, 2001) and EPOC-1 (Sen *et al*, 2004), have also been identified and may play a role in transcriptional regulation during differentiation. It has also been found that ratios of SP1 and SP3 change subtly during differentiation and this change can regulate expression to some degree (Apt *et al*, 1993). Interestingly, linker scanning analysis of the promoter of the closely related virus HPV-31 demonstrated that

many regions are important to the differentiation dependent induction of expression (Ozbun & Meyers, 1999; Pattison *et al*, 1997). This is strong indication that small changes in the activity of many transcription factors may act together to upregulate the p97 promoter during differentiation. Interestingly, the transcriptional changes seen in the studies of factors involved in p97 transcription do not represent the “on/off” switching seen in the late genes, L1 and L2. This is yet another indication that there is an additional level of regulation, and possibly another promoter for the late genes.

Late promoter(s) of HPV-16

It is likely that HPV-16 requires promoters other than the p97 promoter to express its full complement of proteins. The presence of other promoters could explain the tightness of the regulation of products such as L1 and L2. The 5' end of the mRNA that expressed L1 and L2 from wild-type infections has not been fully mapped due to low expression, cell type dependence, and interference from other transcripts. A number of regions from bp 200-700 in HPV-16 have been found to have weak transcriptional activity and may act as late promoters (Fig. 4) (Grassmann *et al*, 1996; Ozbun & Meyers, 1999; Rosensteirne *et al*, 2003). These promoters are named for their transcriptional start sites and include p200, p215, p310, p455, p480, p542, and p670. Interestingly, many of these promoters are not associated with the classical TATAA and CAAT sequences typical of eukaryotic promoters (Fig. 4). The p542 promoter has been shown to be downregulated by the differentiation-dependent factor hSkn-1a, in a closely related virus,

Fig 4. Late promoter(s) of HPV-16



Early and Late promoters are indicated with arrows. Perfect TATAA sequences are indicated with large triangles and close matches are indicated with small triangles. Close matches to the CCAAT sequence are indicated by small triangles.

and is probably downregulated upon differentiation (Ozbun & Meyers, 1999). The p670 promoter has been shown to be upregulated 5-12 fold upon cellular differentiation, and may be responsible for the expression of the late genes (Grassmann *et al*, 1996; Rosensteirne *et al*, 2003). There are no sites for the viral E2 gene in these promoter regions and transcription is likely to be entirely dependent on host cell transcription factors (zur Hausen, 2002; Hebner & Laimins, 2005). The cellular transcription factors that bind these promoters are very poorly defined to date, and remain an interesting target for further research.

1.8 THESIS OUTLINE

Chapter 1 provides a summary of the work described herein as well as a general overview of papillomavirus biology, with particular emphasis on the life cycle of the virus. It has been well established that the life cycle of papillomavirus is intimately linked to the differentiation of the host cell, although few of the molecular mechanisms that control the life cycle have been elucidated. The functions of various viral proteins are also described in Chapter 1, with particular emphasis on the early proteins. The viral E6 and E7 proteins are the major viral oncoproteins and have been linked to cancer progression. The E1 and E2 proteins are involved in control of viral replication and episomal maintenance. The episomal status of the viral genome makes the genome a target for homologous recombination, which is known to occur with high frequency between episomal DNAs. We have exploited this unique aspect of papillomavirus biology in developing a recombination based approach to eliminate papillomavirus infection. The approach utilizes alternative polyadenylation signal usage to regulate the suicide gene Herpes Simplex Virus type I Thymidine Kinase (HSV-TK) on a rAAV delivered cassette. Relevant background information about alternative polyadenylation signal usage, the HSV-TK suicide gene, and rAAV are provided in Chapter 1. As mentioned previously, the complete viral life cycle requires terminal differentiation and yet few of the molecular mechanisms that control this have been characterized. It is known that there are subtle changes in transcription from the major viral promoter (p97) of human papillomavirus type 16 during differentiation. It is thought that one or more late promoters are dramatically upregulated during differentiation and these may be responsible for expression of early as well as late genes. As the viral E2 protein is

dispensable for this upregulation, this is likely due to changes in cellular transcription factor binding during differentiation.

Chapter 2 describes experiments to identify transcription factors with altered binding to the promoters of HPV-16 during cellular differentiation. These factors are likely to control transcription from viral promoters and may be at least partially responsible for the induction of late promoter activity during host cell differentiation. About 35 transcription factors were identified using the methods described, many of which were not previously known to bind HPV-16. Some of these were characterized in greater depth to validate the technique.

Chapter 3 describes the design and testing of a recombination-based approach to eliminate papillomavirus infection. This concept exploits the ability of the episomal papillomavirus genome to undergo homologous recombination and utilizes alternative polyadenylation to regulate the expression of HSV-TK. This approach was effective in eliminating cells infected with Bovine Papillomavirus type I (BPV).

Chapter 4 provides a general discussion of the research described within. Results of the characterization of transcription factors binding to HPV-16 during cellular differentiation are briefly reviewed, and the many unanswered questions that arose from this work are introduced. Many experiments are then described to address these unanswered questions, such as determining the functional significance of these interactions. Next, the rationale behind the design of the TK-expression cassette utilized in the recombination based approach to eliminate papillomavirus infection is explained and supported by detailed case-by-case analysis.

Some of the results of these experiments are reviewed and their implications are discussed. Additionally, future goals in the use of recombination to eliminate viral infections, such as the adaptation of the concept to HPV, are outlined.

Chapter 2

CHARACTERIZATION OF TRANSCRIPTION FACTOR BINDING TO HUMAN PAPILLOMAVIRUS TYPE 16 DNA DURING CELLULAR DIFFERENTIATION

Work described in this section is being published in the Journal of Virology (May 2006, Volume 80) with authors A. Carson and S. A. Khan. A. Carson performed all the experiments described in this section

2.1 INTRODUCTION

Human Papillomaviruses (HPVs) are small, nonenveloped, double-stranded DNA tumor viruses that are associated with over 99% of cervical cancers world-wide (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). More than 100 types of HPVs have been identified to date. Certain mucosatropic HPVs such as types 6 and 11 are rarely seen in cervical cancers, whereas other types, such as 16, 18, 31, and 45 are associated with a much higher risk of cervical malignancy (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). The HPV genome often integrates into the host chromosome at the viral E2 locus in these malignancies, and the resulting cells do not express the viral regulatory protein E2 (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). The loss of E2 activity results in the overexpression of the high-risk HPV-16 E6 and E7 oncoproteins, which promote cell growth by a variety of mechanisms including the inactivation of the functions of the cellular tumor suppressor proteins p53 and pRB (Ciccolini *et al*, 1994; Scheffner *et al*, 1993). Even though HPV-positive cancer cells do not produce infectious virions, understanding the life-cycle of HPVs including the regulation of viral gene expression is fundamental to a better understanding of HPV-associated malignances. Upon infection of the human host, HPVs gain access to the basal epithelium and early genes are transcribed as soon as 8 hours post infection (Ozbun, 2002). The early genes E1 and E2 are required for viral replication (Sverdrup & Khan, 1994), allowing viral DNA to be maintained episomally at low copy numbers in the basal epithelium (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). Virion production requires terminal differentiation of the host cell during which the virus replicates to high copy numbers, produces the capsid proteins L1

and L2, and mature virions are formed (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). The virus that is shed can then reinfect the basal epithelium or spread to new hosts. Many regulated changes in viral gene expression are thought to occur during cellular differentiation, viral DNA amplification and virion production.

The HPV-16 early promoter, p97, is involved in the transcription of the E6, E7 and other viral genes (Dell & Gaston, 2001; Flores & Lambert, 1997; Hebner & Laimins, 2005). Ubiquitous transcription factors (TFs) such as AP-1, NF1, Oct1, and SP1 have been shown to activate transcription from the p97 promoter (Apt *et al*, 1996; Chan *et al*, 1990; Chong *et al*, 1991; Gloss & Bernard, 1990). The YY1 transcription factor has been shown to both activate and repress transcription from this promoter (Don *et al*, 1994). Also, expression from the p97 promoter has been shown to increase slightly during cellular differentiation (Apt *et al*, 1993; Grassmann *et al*, 1996; Yukawa *et al*, 1996). The p670 promoter, which likely corresponds to the late HPV-16 promoter, is also thought to be activated during differentiation resulting in the expression of the late genes L1 and L2 (Grassmann *et al*, 1996; Rosenstierne *et al*, 2003). A number of promoters in HPV-16 have been identified with transcription start sites between positions 200 to 700 that may also play a role in late gene expression. Some of these promoters have been shown to be upregulated during differentiation in HPV-16 (Grassmann *et al*, 1996; Rosenstierne *et al*, 2003), while others are downregulated during differentiation of closely-related viruses (Ozbun & Meyers, 1999). A few differentiation-specific transcription factors such as CDP (Ai *et al*, 1999; O'Conner *et al*, 2000; Pattison *et al*, 1997), and EPOC-1 (Yukawa *et al*, 1996), have also been identified and may play a role in transcriptional regulation of papillomaviruses during differentiation. Other specific

transcription factors such as Sox5 and SRY have previously been shown to regulate transcription of closely-related HPVs (Spink & Laimins, 2005). The ratios of SP1 and SP3 factors are also known to change during differentiation and this in turn may also regulate HPV-16 gene expression (Apt *et al*, 1996). Linker scanning analysis of the early promoter of the high-risk HPV-31 has shown that many TFs and cis-acting sequences are important in differentiation-dependent induction of gene expression (Sen *et al*, 2002; Sen *et al*, 2004). Similarly, a deletion analysis of the late promoter of HPV-31 has revealed many regions and protein factors that are also important in the activation of late promoter during differentiation (Spink & Laimins, 2005). It has also been shown that activation of the late HPV-31 promoter during differentiation is independent of viral genome amplification (Bodily & Meyers, 2005; Spink & Laimins, 2005).

In this study, we have investigated the changes in the levels of cellular transcription factors during differentiation to identify factors that specifically bind to the HPV DNA. For this, we used a Panomics TransSignal array containing 345 binding sites for transcription factors. By competition analysis, we identified the transcription factors that specifically bind to the HPV-16 DNA, including the upstream regulatory region as well as the region containing the late promoter. We found that the binding activity of 36 factors that interact with HPV sequences is altered during differentiation.

We confirmed these results by Western blot analysis and also showed by chromatin immunoprecipitation (ChIP) assays that many of the above transcription factors bind to the HPV-16 DNA *in vivo*. Our data identified several differentially expressed transcription factors that were previously known to bind to HPV-16 DNA during differentiation as well as a number of novel factors that were not previously known to interact with the HPV-16 DNA.

2.2 MATERIALS AND METHODS

Cell line

The HPV-16 cervical cell line 20863 (a clonal population of W12 generated from a low-grade squamous intraepithelial lesion) (Fehrmann & Laimins, 2003) containing episomal HPV-16 DNA was obtained from the laboratories of Drs. Margaret Stanley and Paul Lambert (Medical Research Council, UK and University of Wisconsin, USA, respectively). The cell line was maintained in E medium with 10% FBS in the presence of murine 3T3 J2 fibroblast feeder cells previously treated with mitomycin C (Fehrmann & Laimins, 2003). The 20863 cell line was induced to differentiate by treatment with 1.68% methylcellulose and 10 μ M C8 for 48 hours. Before DNA isolation, fibroblast feeders were removed with versene (phosphate-buffered saline with 0.5 mM EDTA, Gibco BRL).

Panomics array

To investigate the relative binding of human transcription factors (TFs) to their unique consensus sequences, we used the TransSignal protein/DNA combo array with spin column preparation (Cat# MA1215) from Panomics (Redwood City, CA). Array analysis was performed as per manufacturer's instructions using nuclear extracts from 20863 cells before and after differentiation. Nuclear extracts were prepared using the Panomics nuclear extraction kit (Cat# AY2002) as per manufacturer's instructions. Briefly, cells were allowed to swell in a hypotonic buffer to disrupt the plasma membrane. The nuclei were then lysed in a high-salt buffer as described in the Panomics nuclear extraction kit. Protein was quantified by Bradford assay (Bio-Rad Laboratories,

Hercules, CA). Ten micrograms of nuclear extract from 20863 cells isolated before and after differentiation were incubated with 10 μ l TransSignal probe mix (Panomics) for 30 minutes at 15° C. The TransSignal probe mix contains 345 biotin-labeled double-stranded DNA oligonucleotides representing consensus binding sites for approximately 320 transcription factors. This is due to the fact that more than one binding site is included for some TFs, and in some cases more than one TF binds to the same binding site. Twenty microliters of 1X Column Incubation buffer (Panomics) was then added, transferred to a TransSignal spin column, and incubated on ice for 30 minutes. The Spin column was then centrifuged at 7,000 rpm for 30 seconds at 4° C in a refrigerated microfuge and the flow through discarded. Next, 600 μ l of Column Wash buffer was added and incubated on ice for 10 minutes. The Spin column was then centrifuged again at 7,000 rpm for 30 seconds at 4° C and the flow through discarded. The washings with Column Wash buffer was repeated three additional times. Next, 60 μ l of Column Elution buffer was added to the spin column and incubated at room temperature for 5 minutes. Oligonucleotides specifically bound to the TFs present in the nuclear extracts were then eluted by centrifugation for 1 minute at 10,000 rpm in a microfuge. The eluted oligos were then denatured by boiling for 3 minutes followed by a quick chill on ice. These oligos were added to 5 ml of Hybridization buffer (Panomics) and hybridized to the TransSignal Array membrane containing oligonucleotides (representing 345 consensus binding sites for TFs) overnight at 42° C. The blot was then washed at 42° C for 20 minutes with 50 ml of Hybridization Wash I and then washed at 42° C for 20 minutes with 50 ml of Hybridization Wash buffer II. The membrane was then blocked for 15 minutes at room temperature with 20 ml of Blocking buffer (Panomics). Streptavidin-

HRP conjugate was then added to the blocking buffer at a 1:1,000 dilution and incubated for 15 minutes to allow binding to the biotinylated oligonucleotides. The membranes were then washed three times for eight minutes each with Wash buffer (Panomics). The membrane was then incubated with 20 ml Detection buffer at room temperature for 5 minutes. In order to visualize spots, the membrane was overlaid with a mixture of 200 μ l of solution I, 200 μ l of solution II, and 1.6 ml of solution III (all from Panomics) and incubated at room temperature for 5 minutes. The resulting spots were visualized on Kodak X-Omat Blue XB-1 film (Kodak, Rochester, NY) and scanned using a Typhoon scanner. Spots were identified and quantified using the Image Quant software and the data compared using Excel software. In competition experiments, either 20 μ g of pUC19/HPV-16 DNA or 4 μ g of a 1.7-kb *Pst*I fragment of HPV-16 (nt 7005-7816/1-875 that includes all the known promoters) were mixed with 5 μ l of the TransSignal probe mixture followed by incubation with the nuclear extracts made from differentiated 20863 cells. This represents approximately 20-fold molar excess of each competitor over the oligonucleotides in the probe mixture (Panomics, personal communication). The resulting array was then compared to the data obtained using differentiated 20863 nuclear extracts with no competitor HPV-16 DNA. The signals obtained in the presence of the above competitors were then compared with those obtained in the absence of any competitor. Since signal strength varies moderately in each array due to a variety of factors, the total signal strength for each array obtained using the Image Quant software were normalized to each other. On an average, the variation in total signal strength between arrays was less than 1.4-fold. Only factors found to be differentially bound upon

differentiation and competed off by both full-length HPV DNA and HPV promoter regions are reported.

Sequence analysis

In order to further characterize potential TF binding sites identified by the Panomics array, known TF binding sites were obtained from the TRANSFAC database (Wingender *et al*, 2000) and compared to the HPV-16 sequences. The binding sites reported by TRANSFAC facilitated the identification of putative TF binding sites in the HPV-16 regulatory regions and also helped in the design of primer sequences used in the ChIP assays.

Chromatin immunoprecipitation (ChIP) assays

ChIP assays were performed using Upstate Chromatin Immunoprecipitation Assay kit (Upstate, Lake Placid, NY) as per manufacturer's instructions. Briefly, 10^6 cells were trypsinized and cross-linked using formaldehyde, and sonicated with 16 pulses at setting #7 on an Ultrasonics model W-225R sonicator (Plainview, NY). This is expected to generate DNA fragments of an average length of 500-bp. Lysates were then diluted 10-fold in ChIP Dilution Buffer and incubated with salmon sperm DNA/Protein A-agarose slurry for 30 minutes at 4°C with agitation. Cell debris and non-specifically interacting proteins were pelleted by centrifugation for 1 minute at 4°C and the supernatant was collected. An aliquot of this supernatant was used as a "No IP" control in the ChIP assays. Cleared lysates were then incubated overnight along with 20 µl of the appropriate primary antibodies with gentle agitation. All antibodies were purchased from

Santa Cruz Biotechnology (Santa Cruz, CA). Complexes were then immunoprecipitated with Protein-A agarose and washed as per manufacturer's instructions. Proteins were then uncrosslinked from DNA by heat and removed using phenol/chloroform extraction. DNA was then recovered by ethanol-precipitation and subjected to PCR analysis using various HPV-16 specific primer sets. Primer pair I amplifies a region from nt 7,054 to nt 7,221, which is upstream of the established p97 promoter and includes the NFATx site. Primer pair II amplifies a region from nt 7661 to nt 7853, which includes part of the p97 promoter and is close to the NF1 and Pax5 sites. Primer pair III amplifies a region from nt 267 to nt 440, which may be involved in late gene expression and is near two C/EBP α sites and one C-Myb site. Primer pair IV amplifies a region from nt 651-816, which may be involved in late gene expression and is in the vicinity of WT1 and Stat5 binding sites. β -actin primers amplify a region of the β -actin open reading frame with no known transcription factor binding activity. The sequences of the various prime pairs were as follows:

I (HPV 7054-7221): 5'gattgaaggccaaacaaaa3' & 5'caagcacatacaagcacatacaaa3'

II (HPV 7661-7853): 5'taaatcactatgcgccaacg3' & 5'cacacacccatgtgcagttt3'

III (HPV 267-440): 5'gagatgggaatccatagtctg3' & 5'caggacacagtggcttttga3'

IV (HPV 651-816): 5'cagctcagaggaggaggatg3' & 5'gccattaacaggtcttcca3'

β -actin: 5'gccatgtacgttgctatcca3' & 5'aggaaggaaggctggaagag3'

Quantitative PCR analysis of DNA obtained from these ChIPs was performed using the QuantiTect SYBR Green PCR kit from Qiagen (Valencia, CA) as per manufacturer's instructions. As a control, HPV primers set I was used to amplify HPV DNA from the "No IP" samples described above using the QuantiTect SYBR Green PCR kit. The

signals from ChIP amplifications were then normalized to this control HPV amplification to account for unequal cell counts, lysis efficiency, or other differences in sample processing. The annealing temperature for all PCRs was 58° C.

Western blot analysis

Western blot analyses were performed as described (Sambrook *et al*, 1989). Whole cell lysates were prepared from 20863 cells by incubation in a buffer containing 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% Triton X-100 with 2 mM Na₃VO₄ and 50 mM NaF added as phosphatase inhibitors and one pellet of Complete Protease Inhibitor Cocktail (Roche, Mannheim, Germany) per 10 ml. Lysates were cleared by centrifugation for 1 minute at 7,000 rpm in a refrigerated microcentrifuge and quantified by Bradford assay. All antibodies were purchased from Santa Cruz Biotechnology, with the exception of GAPDH, which was purchased from Chemicon (Temecula, CA). Primary rabbit polyclonal antibodies were used at the following dilutions: anti-C/EBP α , 1:1,000 dilution; anti-C-Myb, 1:200 dilution; anti-NF1, 1:1,000 dilution; anti-NFATx (NFATc3), 1:200 dilution; anti-Stat5, 1:100 dilution; anti-WT1, 1:1200 dilution. Primary mouse monoclonal antibodies for Pax5 were used at a dilution 1:200, for GAPDH at a dilution of 1:30,000, and for keratin K10 at a dilution of 1:1,000. All primary antibodies were incubated for two hours in TBS-T (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween 20) with various blocking agents at room temperature. C/EBP α , C-Myb, GAPDH, Stat5, WT1, and keratin K10 primary antibodies were incubated in 5% dry milk. Anti-NFATx antibodies were incubated in 1% dry milk, while anti-NF1 and anti-Pax5 antibodies were incubated in the presence of 5% BSA. Following primary antibody

incubation, blots were successively washed thrice with TBS-T for 5, 10 and 15 minutes with gentle agitation. Secondary anti-rabbit-HRP conjugate was then added at a dilution of 1:3,000 for all blots utilizing rabbit primary antibodies including anti-C/EBP α , anti-C-Myb, anti-GAPDH, anti-NF1, anti-NFATx (NFATc3), anti-Stat5, and anti-WT1. Secondary anti-mouse-HRP conjugate was added at a dilution of 1:5,000 to Pax5, GAPDH, and keratin K10 blots since they utilized mouse primary antibodies. The blots were incubated for one hour at room temperature in TBS-T containing the same blocking agent as that used with the primary antibodies. Blots were again successively washed for 5, 10, and 15 minutes and proteins were visualized using the ECL plus Western Blotting Detection System (Amersham, Buckinghamshire, UK) as per manufacturer's instructions.

2.3 RESULTS

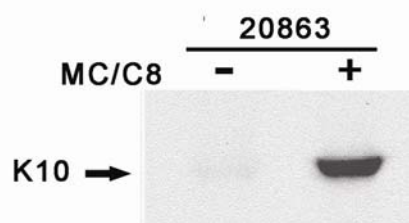
The 20863 cell line containing HPV-16 can be differentiated by methylcellulose and C8 treatment

Prior to their use in our experiments, we confirmed the presence of episomal HPV-16 DNA in the W12 clone 20863 by isolation of the Hirt fraction (Hirt, 1967)) followed by Southern blot analysis (data not shown). To differentiate 20863 cells, we used a combination of methylcellulose and C8 treatment. Methyl cellulose is an established method to differentiate keratinocytes ((Fehrmann & Laimins, 2003). C8 is a synthetic diacylglycerol that activates the protein kinase C pathway to induce a more complete differentiation (Sen *et al*, 2002; Sen *et al*, 2004). These treatments have been used together to induce differentiation in HPV-positive cells (Sen *et al*, 2002; Sen *et al*, 2004). Treatment of the 20863 cell line with methylcellulose and C8 resulted in induction of the differentiation marker keratin K10 as shown by Western blot analysis (Fig. 5), demonstrating that differentiation has been induced.

Identification of Transcription Factors with Altered Expression/Activity during Differentiation that bind to HPV-16 DNA

We utilized a Panomics TranSignal array to compare the binding of TFs present in nuclear extracts of undifferentiated and differentiated 20863 cells to oligonucleotides representing 345 different consensus sequences. This analysis identified binding increases of over 2-fold for 64 consensus sequences and a decreases of over 2-fold for 86 sequences (data not shown). This corresponds to changes in the binding activity of 148

Fig 5. 20863 cells express the differentiation marker keratin K10 upon treatment with methylcellulose and C8



Cell lysates were prepared from treated or untreated cells and subjected to western blot analysis using anti-keratin K10 antibodies. K10, keratin 10; MC, methylcellulose.

different TFs upon differentiation (the ISRE and CBF/NF-Y factors have two binding sites each on the array).

In order to identify TFs that show differences in their activity during differentiation and that also bind to HPV-16 sequences, a series of competition assays were performed. In one experiment, a pUC19 plasmid containing the full-length HPV-16 genome (Sverdrup & Khan, 1994) was added in a 20-fold molar excess over the target oligonucleotides during incubation with the nuclear extracts. The binding of TFs to the exogenously added HPV-16 DNA is expected to compete with their binding to the biotinylated target oligonucleotides. This would result in a loss/reduction in signal obtained upon hybridization of the labeled oligonucleotides to the array (see Materials and Methods). The binding of 48 transcription factors to the oligonucleotides was competed off (greater than 4-fold reduction in probe binding to the array). These data suggested that upon differentiation the binding of 38 TFs to cloned HPV-16 is increased while that of 9 TFs is decreased. We also used a 1.7-kb *Pst*I fragment of HPV-16 (nt 7005-7816/1-875) which includes both the p97 and late promoters of HPV-16 as a competitor in the above assays. The data obtained from the array showed that the above fragment competed for the binding of 31 TFs that showed increased binding and 5 that showed decreased binding upon differentiation to the probe oligonucleotides. The cloned pUC19/HPV-16 DNA contains both specific and “non-specific” sequences, and the “non-specific” sequences also serve as a negative control in these experiments. Note that most of the TFs competed off by pUC19/HPV-16 DNA were also competed off by the known HPV-16 regulatory regions present in the 1.7-kb *Pst*I fragment. This provides a validation of the technique and shows that most of the TFs specifically interact with the

TABLE 1. Changes in transcription factor binding to HPV-16 promoters during cellular differentiation

TF	Fold Increase*	TF	Fold Decrease
AP-2	6.39	AP-1	2.93
ATF2	+++	MyoG	6.76
C/EBP α	7.61	GRE	14.6
CACC	+++	Pbx1	2.04
CCAAT	+++	TREF1, 2	33.0
C-Myb	+++		
CP1	+++		
CPIB	+++		
EBP40, 45	10.1		
ETF	+++		
GAG	5.87		
GBF1, 2, 3, HY5	+++		
GKLF	3.03		
H4TF-1	+++		
HIF-1	51.4		
KTP	+++		
Lactoferrin BP	+++		
LCR-F1	+++		
MT-box	5.75		
MTF	+++		
NF1	+++		
NFAtx	+++		
NF-kB	3.83		
NF-Y	+++		
Pax5	+++		
PCF	2.03		
PEBP	+++		
RFX1, 2, 3	20.3		
RIPE3a	+++		
RORE	2.56		
Snail	+++		
Stat5	3.83		
TFE3L	+++		
WT1	25.3		
YY1	2.38		

*-When TF binding was not detectable in undifferentiated cells but present upon differentiation, an infinity value was assigned by the software, these are labeled (+++).

HPV-16 regulatory regions. Based on the above experiments, the TFs that specifically bind to a region of HPV-16 that contains known promoter regions are listed in Table 1.

The HPV-16 promoter regions have identifiable consensus sequences for novel TFs whose expression is altered during differentiation

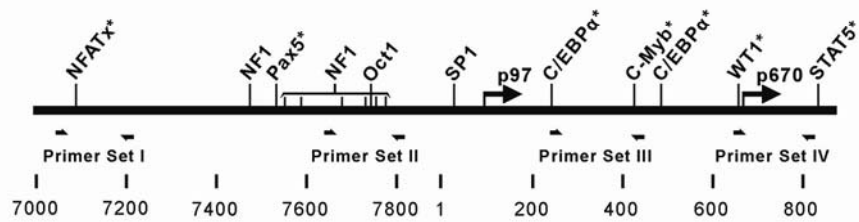
Many TFs identified in the above experiments were previously not known to bind to the HPV-16 promoters and may represent novel interactions. Therefore, we analyzed the sequence of the *Pst*I fragment of HPV-16 (nt 7005-7816/1-875) for the presence of consensus binding sites for known human TFs in the TRANSFAC database (Wingender *et al*, 2000). A number of binding sites for both TFs that are known to interact with HPV-16 DNA as well as novel factors that represent previously unknown interactions with the viral DNA were identified in this region, including several NF1 binding sites, two sites for C/EBP α , and single sites for C-Myb, Stat5, Pax5, WT1, and NFATx (Fig. 6).

The levels of several transcription factors that bind to HPV-16 sequences are increased during differentiation

The Panomics TranSignal array detects changes in the binding of known transcription factors to their binding sites. While this would generally indicate altered expression of TFs during differentiation, it may also be due to changes in the activity of TFs (for example, due to protein modifications). Therefore, we also carried out Western blot analysis for a few representative TFs to determine if the increase in their binding

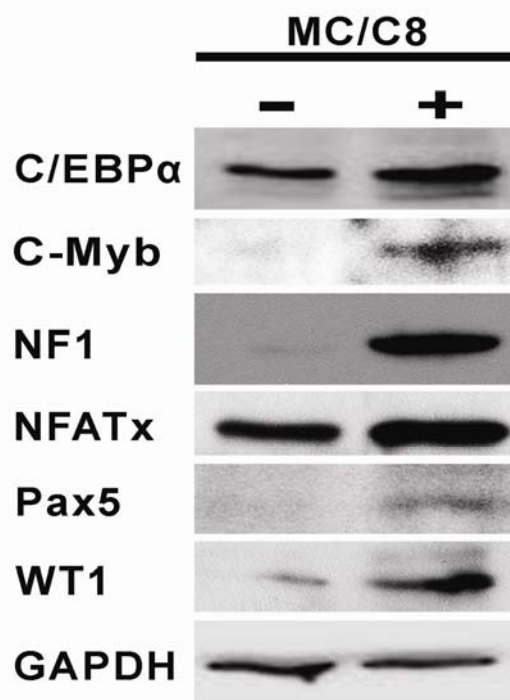
activity during differentiation as determined by using the Panomics array is correlated with increased protein levels. Fig. 7 is a representation of three independent experiments and shows that the levels of C-Myb, NF1, Pax5 and WT1 were dramatically increased upon differentiation, while the levels of C/EBP α , and NFATx showed only a modest increase. These results showed that the expression of several TFs that bind to HPV-16 promoter/regulatory regions is upregulated during differentiation.

Fig 6. Putative TF binding sites in the HPV-16 promoters and upstream regions



TFs found to bind to HPV-16 sequences in this study are indicated with a star (*). The binding sites for some TFs previously shown to bind to HPV-16 are also indicated. The p97 and p670 promoters are indicated along with the location of the primers used for PCR and QPCR shown in Fig. 4.

Fig 7. Western blot analysis of TFs in differentiated and undifferentiated cells

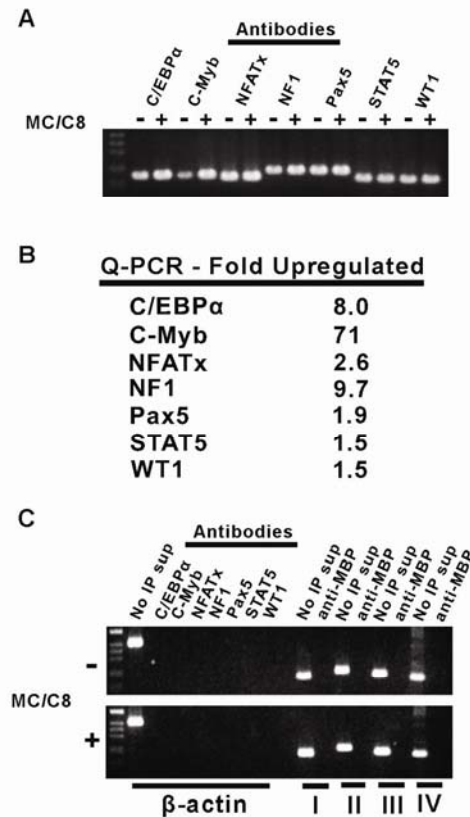


Whole-cell lysates were subject to western blot analysis using the appropriate antibodies as described in Materials and methods. MC, methylcellulose.

Binding of transcription factors to the HPV-16 DNA *in vivo*

Chromatin immunoprecipitations (ChIPs) experiments were carried out to determine whether some of the TFs identified in the above screen bind to the HPV-16 DNA *in vivo*. For this, proteins bound to DNA in the 20863 cell line with or without differentiation were cross-linked with formaldehyde and specific DNA-protein complexes isolated using antibodies against selected transcription factors as described in Materials and Methods. The DNA bound to proteins was recovered and the bound HPV sequences were identified by standard PCR analysis as shown in Fig. 8A. In order to quantify the amount of HPV DNA present, the above samples were also analyzed by quantitative PCR using primer pairs near the predicted binding sites for the above transcription factors (see Fig. 6). The relative amount of bound DNA is expected to represent the relative levels of specific TFs bound to the viral DNA in differentiated vs. undifferentiated cells. The data, which are averages of three experiments, showed that C/EBP α , C-Myb, and NF1 bound to the HPV-16 DNA in differentiated cells to substantially higher levels than in undifferentiated cells with increases of 8-, 71-, and 9.7-fold, respectively (Fig. 8). NFATx, Pax5, Stat5, and WT1 showed a more moderate increase in binding to HPV sequences upon differentiation (2.6-, 1.9-, 1.5-, and 1.5-fold, respectively). In control experiments, the use of template DNA immunoprecipitated

Fig 8. Chromatin immunoprecipitation (ChIP) and quantitative PCR analysis of TF binding



(A) Non-quantitative PCR analysis of HPV-16 DNA recovered from immunoprecipitates from 20863 cell lysates using the indicated antibodies. (B) Relative increase in the binding of various TFs to HPV-16 DNA in 20863 cells upon differentiation. (C) Control experiments demonstrating the specificity of the antibodies used in the ChIP assays. In the left half of each panel, a region of the β -actin open reading frame with no known TF binding activity was PCR amplified from ChIP samples using the indicated antibodies. The right half of each panel shows that the HPV-16 PCR primer pairs do not amplify DNA from material obtained using control antibodies. No IP sup, supernatant prior to immunoprecipitation.

using antibodies specific to various TFs failed to amplify the β -actin gene (Fig. 8C). In another control, we used anti-MBP antibodies in the above experiments and the “mock-precipitated” DNA was used as the template for PCR amplification using HPV-16 primer pairs I, II, III and IV described above. No amplified DNA was obtained (Fig. 8C). The above ChIP assays identified TFs that bind to HPV-16 regulatory regions *in vivo* and also revealed the TFs whose binding to the viral DNA is increased upon differentiation.

2.4 DISCUSSION

We have used the W12 clone 20863 to identify changes in TF binding to HPV-16 promoter/regulatory regions during host cell differentiation using the Panomics TransSignal array. The 20863 cell line harbors HPV-16 episomally and can be induced to differentiate (Fig. 5). Our analysis identified changes (≥ 2 -fold difference) in the binding of TFs to 150 consensus sequences during differentiation out of the 345 sequences present on the array. These changes in TF binding may either represent their altered expression, activity, or both. Competition experiments using the Panomics TransSignal arrays revealed 38 TFs present in differentiated 20863 cells with increased binding to the pUC19/HPV-16 DNA, while 9 TFs in differentiated cells showed a decrease in binding to this DNA (Table 1). Since TF consensus sequences may also be present randomly in the HPV-16 genome or the vector pUC19 sequence and may have no role in transcription, we also used a 1.7-kb *Pst*I fragment of HPV-16 (nt 7005-7816/1-875) which contains all the known viral promoter/regulatory elements as competitor in the above experiments. These results identified 31 TFs with increased binding to the above 1.7-kb region while 5 TFs showed reduced binding upon differentiation (Table 1). These TFs were a subset of those identified when the pUC19/HPV-16 DNA was used as the competitor. We then analyzed the above 1.7-kb region for the presence of putative binding sites for known TFs using the TRANSFAC database (Wingender *et al*, 2000). A majority, but not all of the TFs identified by the Panomics array have identifiable binding sites in this region (Fig. 2 and data not shown). The putative binding sites of many of the above TFs are located in regions associated with late promoters and these factors may play a role in controlling late gene expression. These poorly characterized promoters

(with transcription start sites located between nt 200 and 700 of the HPV-16 genome) are thought to be dramatically upregulated during host cell differentiation as compared to the slight upregulation of the p97 promoter (Grassmann *et al*, 1996; Rosenstierne *et al*, 2003). A plausible hypothesis is that this upregulation of the late promoters is due to increased binding of stimulatory TFs (including those identified in this study) during differentiation. It is interesting, therefore, that while approximately similar number of TFs show increased and decreased binding to their consensus sequences during differentiation (65 and 85, respectively), the vast majority of TFs that interact with the regulatory regions of HPV-16 show increased binding activity during differentiation (Table 1). Many of these factors are known to stimulate transcription and may play a role in the upregulation of late promoters during the viral life cycle.

Several of the TFs found to be upregulated during differentiation in this study such as C/EBP α and NF1 have previously been shown to affect HPV-16 transcription (Apt *et al*, 1993; Chong *et al*, 1991; Dong *et al*, 1994; Hadaschik *et al*, 2003). Such known factors provide a validation of the approach we have used in this study which also identified several TFs not previously implicated in the regulation of HPV gene expression. Our data also expand the knowledge of previously known interactions between HPVs and TFs. For C/EBP α , this represents the first in vivo evidence that this factor binds to the HPV-16 DNA and the first demonstration that this binding and protein expression increases during differentiation. The dramatically increased DNA binding activity of NF1 during differentiation is most likely due to a significant increase in its expression as shown by Western blot analysis (Fig. 7). NF1 has been shown to upregulate the p97 promoter through its multiple binding sites located in this region (Apt

et al, 1993; Chong *et al*, 1991). Many other TFs such as Stat5 have been loosely associated with HPV infection or HPV-related malignancies but were not known to be directly involved in transcription of the viral genome (Chang & Laimins, 2000; Havard *et al*, 2005; Rocha-Zavaleta *et al*, 2004; Nees *et al*, 2001). The activity of Stat5 protein was consistently increased in both oligonucleotide binding as determined by the Panomics array as well as in vivo binding to the HPV-16 genome as assayed by ChIP (Table 1 & Fig. 8). Our results further support a role for Stat5 in HPV gene expression during viral maturation or virus-induced carcinogenesis.

We have identified several TFs in this study and provide the first in vivo evidence for their binding to HPV-16 DNA as well as changes in their binding to the viral genome during differentiation. For example, binding of C-Myb, Pax5, NFATx and WT1 to the HPV-16 regulatory region represents novel interactions. The Wilm's tumor gene product (WT1) is known to be associated with tumor progression (Lee & Haber, 2001), and its interaction with the HPV-16 promoter regions could represent an additional mechanism by which this factor may promote carcinogenesis. TFs such as C-Myb, NFATx, and Pax5 are known to play important roles in development and/or differentiation (Serfling *et al*, 2000; Hirokawa *et al*, 2003; Rushton *et al*, 2003). Expression of the C-Myb and Pax5 protein was not detectable in the 20863 epithelial cells before differentiation, but was induced during differentiation as shown by Western blot analysis (Fig. 7). While Western blot analysis showed a limited increase in the levels of NFATx during differentiation (Fig. 7), its DNA binding activity was significantly increased as determined by the Panomics array and ChIP assays (Table 1 & Fig. 8). This could be due to changes in nuclear localization since the NFATx activity is known to be regulated in

this manner (Serfling *et al*, 2000). It is likely that HPV-16 and possibly other HPVs may have adapted to utilize many cellular transcription factors involved in differentiation and development to regulate their gene expression during host cell differentiation in epithelial cells.

The relative changes in TF levels or activity during differentiation as identified by the Panomics array, ChIP assay, and Western blot analysis were variable. These variations may represent differences in the sensitivity of the assays, increased activity resulting from protein modifications and/or subcellular localization, or due to some unknown factors. Since we have validated our data by carrying out detailed analysis of a select number of TFs identified using the Panomics array, it increases the confidence that other TFs identified using this array may also interact with HPV-16 sequences and their expression/activity may also be altered during differentiation.

Taken together, our results as well as previous studies by other investigators suggest that HPVs may have evolved to utilize many changes in transcription factor binding and expression during differentiation to achieve tight regulation of the late viral genes. Many such factors are involved in differentiation, development and/or carcinogenesis and should be targets for further study. In the future, we plan to confirm interactions of additional TFs with HPVs and determine the functional significance of many of the novel transcription factors in HPV gene expression and HPV-associated disease.

Chapter 3

A DNA RECOMBINATION-BASED APPROACH TO ELIMINATE PAPILLOMAVIRUS INFECTION

Work described in this chapter was published in *Gene Therapy* (2005, Volume 12, pages 534-540) with authors A. Carson, Z. Wang, X. Xiao, and S. A. Khan. Z. Wang and X. Xiao provided plasmids into which the expression cassettes were inserted, prepared the recombinant AAV used in this study and took part in many helpful discussions. A.

Carson performed all the other experiments described in this section

3.1 INTRODUCTION

Papillomaviruses (PVs) are small, nonenveloped, double-stranded DNA tumor viruses (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). During productive infection, the virus gains access to the basal epithelium where the viral DNA replicates to a low level of approximately 50 copies per infected cell (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Virion production requires terminal differentiation of the host cell during which the virus replicates to high levels, produces the capsid proteins L1 and L2, and mature virions are formed (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). The shed virions then reinfect the basal epithelium or they can infect new hosts. PVs infect many different species of animals in nature, and to date more than 100 types of human papillomaviruses (HPVs) have been identified (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Certain mucosatropic HPVs such as types 6 and 11 cause low-grade cervical lesions, while others such as HPV-16, HPV-18, HPV-31, and HPV-45 are strongly associated with cervical cancer and other anogenital cancers (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). The HPV genome often integrates into the host chromosome at the viral E2 locus in these malignancies, and the resulting cells do not express the viral regulatory protein E2 (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). The loss of E2 activity results in the overexpression of the viral E6 and E7 oncoproteins, which inactivate the functions of the cellular tumor suppressor proteins p53 and RB, respectively (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman &

Laimins, 2003; zur Hausen, 2002). The E6 and E7 proteins also target several other cellular pathways that contributes to their oncogenic potential (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Malignancies associated with HPV infection usually occur several years after the primary infection and the oncogenic potential of HPVs is greatly increased by its ability to cause persistent infections. Although low-risk HPVs such as types 6 and 11 are rarely associated with cancer, they often cause painful long-term infections (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Also, cutaneous HPVs such as types 1 and 2 can cause benign lesions that can be painful or cosmetically unacceptable and require medical treatment (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002).

Untreated, the above types of lesions may persist for decades without clearance or regression, and surgical treatment does not prevent recurrence. (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). At present, no effective treatment is available to target and eliminate HPV from persistently infected cells. During the persistent state, large population of host cells harbor episomal HPV DNA. This allows a potential opportunity to disrupt the episomal HPV genome by an approach involving homologous recombination. In contrast to the chromosome, homologous recombination between episomal DNA elements occurs more frequently. Intermolecular homologous recombination between related plasmids cotransfected into immortalized cells occurs in roughly 0.2–1% of recovered plasmids (Ayers *et al*, 1985; Thyagarajan *et al*, 1996). These frequencies have been shown to be even higher in a

bovine papillomavirus (BPV)-based mammalian plasmids over extended time periods (Kitamura *et al*, 1990).

In this study, we have developed a DNA recombination-based approach for the conditional expression of the suicide HSV-1 thymidine kinase (TK) gene. The gene cassette is designed to express TK, through homologous recombination, only in cells that contain BPV DNA. Transfection of the DNA cassette resulted in the expression of TK in cell lines containing the BPV DNA, whereas little or no expression was observed in cells lacking the BPV sequences. We also utilized an adeno-associated virus (AAV)-TK cassette for the specific expression of the HSV-1 TK gene in BPV-1-containing cells. In combination with ganciclovir (GCV) therapy, this novel approach in gene therapy may allow elimination of PV infections by targeting the viral DNA specifically in infected cells.

3.2 MATERIALS AND METHODS

Oligonucleotides

The following primers were used for the amplification of DNA for cloning:

TK-forward:GGAGATCTATGCCCACGCTACTGCGG

TK-reverse:GGTCTAGAGGACGCGTTCAGTTAGCCTCCCCAC

Polyoma-forward:ACGCGTGCTATGTTGACCGCTTTG

Polyoma-reverse:TCTAGACTCCTACAGTTTGTGTTTGTG

BPV-forward:CCTTTAGACACCCTTGAACTC

BPV-reverse:GGAATTCGACCACAGCTACATGTCAGG

3S³-forward:GGGGATCCTGCAGGTAAGTCCGATGACACTACTACTACACC

3S³-reverse:GACCGTCTATATAAACCCGC

3S5A-forward:GGCTGATTATGATCCTCTAGAGTCGCAG

3S5A-reverse:GGTGATCAGGTTTTCCCAGTCACGACGTTGTAAAAC

3S5B-forward:GGTGATCACTGCGACTCTAGAGGATCATAATCAGCC

3S5B-reverse:GGGGTGATCAGAAGCGGAAGAGTCTAGATCCTACAGTTTG

The following oligonucleotides were used to create 3' ss II and the FLAG epitopes:

3' ss II-top:CTAGCCTAACCATGTTTCATGCCTTCTTCTTTTCCTACAGA

3' ss II-bottom:GATCTCTGTAGGAAAAAGAAGAAGGCATGAACATGGTTAGG

FLAG-top:AATTACCACCATGGACTACAAAGACGACGACGACAAAG

FLAG-bottom:AATTCTTTGTCGTCGTCGTCTTTGTAGTCCATGGTGTT

DNA constructs

A truncated version of the TK ORF (Salomon *et al*, 1995) was PCR amplified using primers TK-forward and TK-reverse. The ORF was then isolated as a *Bgl*II/*Mlu*I fragment. The weak PV poly(A) site was PCR amplified from the wild-type A2 strain of PV using primers polyoma-forward and polyoma-reverse. This poly(A) site was then isolated as a *Mlu*I/*Xba*I fragment. These two fragments were then ligated into *Bgl*II/*Xba*I-digested plasmid pSG5 (Stratagene) in a three-way ligation to create pSTP. The BPV-1 genome from nt 4501 to base 6647 containing portions of the L1 and L2 ORFs was PCR amplified using primers BPV-forward and BPV-reverse. This region was then isolated as an *Eco*RI fragment and cloned into the *Mfe*I site of pSTP to create plasmid S2.1TP. The SV40 poly(A) site of pSG5 was isolated as a 160-bp *Bam*HI/*Xba*I fragment, the ends were filled-in with Klenow fragment of DNA polymerase I and ligated into the filled-in *Nhe*I/*Cla*I site within the 2.1-kb BPV sequence described above. The *Nhe*I/*Cla*I digestion within the BPV-1 sequence resulted in the loss of 46 bp from the BPV sequence between the BPV^a and BPV^b regions. An *Eco*RI fragment containing two SV40 poly(A) sites in the forward orientation along with the adjacent BPV^a and BPV^b sequences was then isolated and cloned into the *Mfe*I site of pSTP to create SS²TP. The SV40 poly(A) trimer was released from plasmid pUCA15 (Batt *et al*, 1994) via *Hpa*II/*Bgl*II digestion. A double-stranded oligonucleotide with *Bgl*II and *Nhe*I ends containing the 3' ss II sequence along with the above SV40 poly(A) trimer was ligated into the *Nhe*I/*Cla*I site of the 2.1-kb BPV fragment described above in a three-fragment reaction. These elements were then isolated as an *Eco*RI fragment and cloned into the *Mfe*I site of pSTP to create plasmid 3S³. A region of 3S³ was PCR amplified using primers 3S³-forward and 3S³-

reverse to introduce the 5' ss II upstream of the BPV^b region. The amplified product was then isolated as a *Bam*HI fragment and ligated into *Bam*HI-digested 3S³. This resulted in the addition of the 5' ss II to 3S³, generating plasmid 3S³5. A 378-bp region of the 3S³5 plasmid was deleted from the BPV^a region by digestion with *Nhe*I and *Afl*III, filling in the ends with Klenow, and self-ligation to create 3S³5-del. A synthetic FLAG epitope was created by hybridizing oligonucleotides FLAG-top and FLAG-bottom with *Eco*RI ends, and ligating the DNA upstream of the TK ORF into the *Eco*RI site of 3S³5-del to create 3S³5-dF. The same FLAG epitope was also inserted into the *Eco*RI site of S2.1TP to create S2.1TP-F. We also generated the AAV-3S5 vector for delivery of the TK cassette into murine cells. This involved cloning of a TK cassette similar to 3S³5-dF, except containing a single SV40 poly(A) site in place of a trimer, into the AAV vector pXX. For this, two regions of 3S³5-dF cassette were amplified, one containing the SV40 promoter and the BPV^a region (using primers 3S5A-forward and 3S5A-reverse) and the second containing one copy of the SV40 poly(A) signal, BPV^b region, the TK ORF and the polyomavirus poly(A) signal (using the primers 3S5B-forward and 3S5B-reverse). The first region was then isolated as a *Bcl*II/*Bgl*II fragment and the second as a *Bcl*II fragment. These two fragments were then simultaneously ligated into the *Bgl*II sites of plasmid pXX containing the AAV inverted terminal repeats to generate the AAV-TK plasmid pXX-3S5.

Virus production

The AAV used was serotype 2. The recombinant viral stocks were produced according to the adenovirus-free, three plasmid cotransfection method (Zolotukhin *et al*, 1999). Briefly, the AAV-TK vector plasmid pXX-3S5 was cotransfected into human 293 cells with the helper plasmids pXX2 (Zolotukhin *et al*, 1999) supplying *rep-cap* function and pXX6, a truncated adenovirus plasmid supplying E2a, E4, and VA RNA. Cells were harvested 48 h after transfection, and then frozen and thawed four times. The viral particles were purified from the crude lysate using an FPLC-Heparin column (Zolotukhin *et al*, 1999). Virus particle numbers were determined by quantitative DNA dot-blot analysis (Sambrook *et al*, 1989).

Transfections

COS-1 cells were transfected with the lipofectamine reagent (Gibco). Transfections were performed according to the supplier's instructions in 60 mm dishes seeded with $2-3 \times 10^6$ cells, 3 μ g total DNA, and 30 μ l lipofectamine reagent per transfection.

Infections

BPV-137 or C-127 cells were plated 1 day prior to infection at a density of $2-3 \times 10^5$ cells per 60 mm dish. The next morning, cells were treated with 3 mM hydroxyurea for 2-3 h in complete DMEM media. Following this treatment, cells were washed and complete media containing 5×10^9 viral particles was added. Infections were allowed to proceed overnight, the cells were again washed, and complete media were added.

Western blots

Western blots were performed as described (Sambrook *et al*, 1989). Primary rabbit polyclonal anti-TK antibodies were purchased from William Summers at Yale University and used at a 1:400 dilution. Primary anti-FLAG M2 monoclonal antibodies were purchased from Sigma and used at a 1:1000 dilution.

RNA isolation and northern blots

RNA was isolated using the ULTRASPEC RNA Isolation Kit from Biotecx according to the manufacturer's instructions. Northern blot analysis was performed by standard methods (Sambrook *et al*, 1989). The radiolabeled probes were prepared by random primer labeling of restriction enzyme fragments or whole plasmids (Sambrook *et al*, 1989). The blots were stripped by boiling in 25 mM phosphate buffer (pH 7.2) with 5 mM EDTA and 1% SDS.

GCV treatment

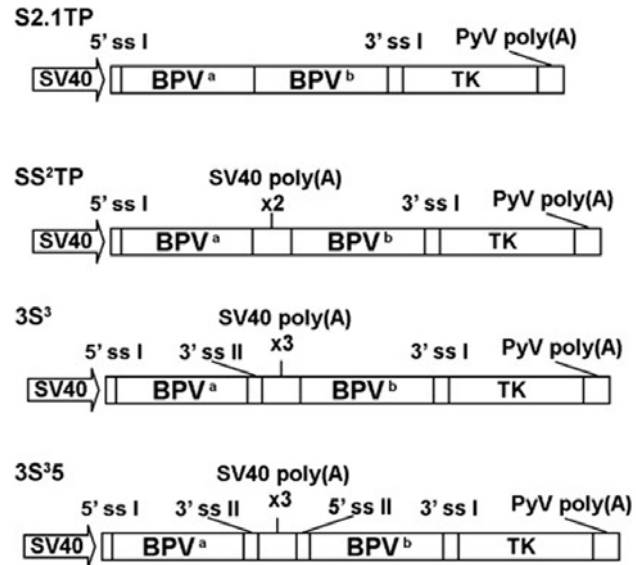
BPV-137 cells were plated 1 day prior to infection at a density of $2-3 \times 10^5$ cells per 60 mm dish (in triplicate). The next morning, cells were treated with 3 mM hydroxyurea for 2–3 h in complete DMEM media. Following this treatment, cells were washed and complete DMEM medium containing 5×10^9 viral particles was added. Infections were allowed to proceed overnight, the cells were again washed, and complete DMEM medium was added (Mizuno *et al*, 1998). GCV was added to a concentration of 50 μ M and selection was allowed to proceed for 4 days (Fukui *et al*, 2001).

3.3 RESULTS

Development of Recombination-Dependent TK Expression Cassettes

We developed a strategy for the expression of the HSV-1 TK gene conditional upon its recombination with PV sequences. The desirable features of this gene cassette are that it must not express TK, but in the presence of PV sequences in the cell, it should undergo homologous recombination, leading to the expression of the TK gene. The DNA sequence encoding the amino-terminal region of TK has been shown to act as a weak promoter in some tissues (Salomon *et al*, 1995). As TK expression must be tightly regulated in this system, we utilized an amino-terminal-deleted version of TK in which this promoter is deleted. This TK deletion retains the ability to phosphorylate GCV (Salomon *et al*, 1995). We developed several TK expression cassettes as shown in Fig. 9. In preliminary studies, we first utilized a strategy involving the use of alternative polyadenylation (poly(A)) sites to restrict TK expression prior to recombination. We generated the SS²TP expression cassette to take advantage of the known criteria for poly(A) site selection (for a review see Edwards-Gilbert, Veraldi & Milcarek, 1997). In this expression cassette, an intron containing nt 4501–6647 of the BPV genome consisting of parts of the L2 and L1 ORFs was inserted upstream of the TK open-reading frame (ORF) and the weak late poly(A) site of polyomavirus (Batt *et al*, 1994). Two poly(A) regions that contain both the early and late poly(A) sites of SV40 were then inserted into the middle of the above BPV sequence such that the upstream BPV^a region consisted of 926 bp, while the downstream BPV^b region contained 1169 bp (Fig. 9). We anticipated that transcripts from this expression cassette would primarily

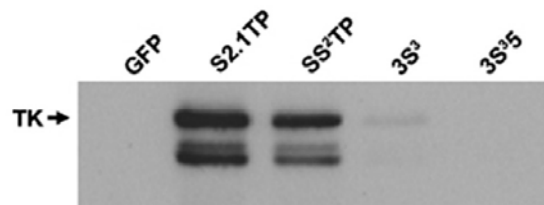
Fig 9. Diagram of various TK constructs used in this study



TK, thymidine kinase; ss, splice site; PyV, polyomavirus; SV40, Simian virus 40; BPV^a and BPV^b sequences correspond to the L1 and L2 regions of BPV-1. See text for details. The large arrow on the left indicates the SV40 early promoter.

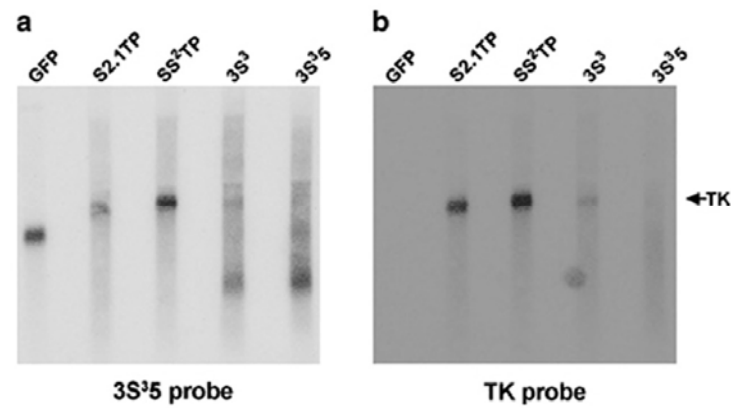
terminate at the strong upstream SV40 poly(A) sites. Western blot analysis using protein extracts from COS-1 cells transfected with SS²TP showed that TK expression from this plasmid was nearly as high as in cells transfected with the positive control plasmid S2.1TP that is expected to express high levels of TK (Fig. 10). A faster-migrating band in the Western blot presumably corresponds to a breakdown product of TK. A plasmid containing the GFP gene was used as a negative control and it did not express any TK (Fig. 10). In order to further characterize expression from the SS²TP plasmid, a series of Northern blot analyses were performed using total RNA isolated from COS-1 cells transfected with various plasmids. When the Northern blot was probed with the 3S³5 plasmid DNA, a single band of approximately 1400 nt was observed in cells transfected with SS²TP (Fig. 11a). COS-1 cells transfected with the control TK expression plasmid S2.1TP also contained a single transcript of approximately 1400 nt (Fig. 11a). The transcript present in the GFP plasmid presumably corresponds to the GFP transcript, which also contains 5' and 3' regions that are common to all the plasmids. The blot was then stripped and reprobed with the TK ORF to identify transcripts containing TK sequences (Fig. 11b). The TK probe hybridized to the same RNA species in cells transfected with S2.1TP and SS²TP plasmids that were detected using the 3S³5 probe, demonstrating that this transcript corresponds to the TK mRNA (Fig. 11b). No specific band was detectable when the same blot was reprobed with BPV sequences (data not shown). Taken together, the results of Northern blot analyses indicated that only one transcript was produced by the S2.1TP and SS²TP expression cassettes, which contains the TK ORF but no BPV sequences. Furthermore, the approximate size of this transcript

Fig 10. Western blot analysis of TK expression



COS-1 cells were transfected with the indicated plasmids and protein lysates were subjected to Western blot analysis using anti-TK polyclonal antibodies. The position of the TK band is indicated by an arrow. The faster-migrating band presumably corresponds to a TK breakdown product.

**Fig 11. Northern blot analysis
of COS-1 cells transfected with the
various TK plasmids**



Total RNA was harvested 3 days post-transfection and subjected to Northern blot analysis. The Northern blots were probed either with the entire 3S35 expression cassette (a) or with a TK-specific probe (b).

is consistent with termination of transcription at the polyomavirus poly(A) site, and little or no termination at the SV40 poly(A) sites.

We hypothesized that TK expression by the SS²TP plasmid may be due to the fact that the SV40 poly(A) sites closely follow a 5' splice site (5' ss I) in this plasmid, and the unspliced transcript may not be efficiently terminated/polyadenylated. To eliminate this possibility, we inserted a 3' splice site (3' ss II) upstream of the SV40 poly(A) sites. We also replaced the SV40 poly(A) dimer with an SV40 poly(A) trimer to possibly increase the relative strength of the polyadenylation signal. This trimer contains three head to tail copies of a 237 bp segment of SV40 containing both the early and late poly(A) signals (Maxwell *et al*, 1989). The resulting construct (3S³ – Fig. 9) showed greatly reduced TK expression as assayed by Western blot analysis (Fig. 10). However, TK expression was not completely eliminated in the presence of this plasmid (Fig. 10 and data not shown). Northern blot analysis of these transfectants indicated that two transcripts of approximately 1400 and 400 nt were produced by this expression cassette (Fig. 11a). The larger transcript hybridized to the TK ORF but not to the BPV sequences (Fig. 11b and data not shown). The smaller transcript contained neither TK nor BPV sequences, consistent with it being an RNA species resulting from the usage of the SV40 poly(A) sites that will contain only 5' and 3' untranslated regions (Fig. 11b and data not shown). Together, these results suggest that the upstream SV40 poly(A) site is used predominantly in this construct, but the downstream polyoma poly(A) site is also used to a limited extent resulting in low-level TK expression.

It is likely that low-level TK expression from the 3S³ plasmid is due to alternative splicing from the 5' ss I to the 3' ss I (Fig. 10) by a mechanism similar to exon skipping. As the skipped exon in this system had no 5' splice site following the 3' ss II and the SV40 poly(A) signals, we introduced another 5' splice site (5' ss II) at this position, generating the 3S³5 cassette (Fig. 9). When this plasmid was transfected into COS-1 cells, TK expression was undetectable by Western blot analysis (Fig. 10). Northern blot analysis revealed that the 400 nt transcript first seen with 3S³ is more abundant with 3S³5 (Fig. 11a). No transcripts complementary to the TK or BPV L1 probes were observed upon transfection of 3S³5 (Fig. 11b and data not shown), demonstrating that the upstream SV40 poly(A) sites were exclusively used in this construct.

The ability of homologous recombination to regulate TK expression

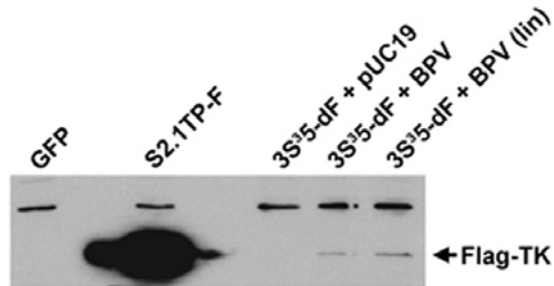
The 3S³5 plasmid was designed to express TK when this construct is introduced into PV-infected cells but not in uninfected cells. The regulation of TK in this system is expected to depend on homologous recombination between segments of the endogenous PV genomes and PV L1/L2 sequences present in the 3S³5 construct. To assess this possibility, we modified 3S³5 to create the plasmid 3S³5-dF. First, a sequence encoding the FLAG epitope was added upstream of the TK ORF to generate a FLAG-TK fusion to increase the sensitivity of the Western blots. Second, the overall size of the expression cassette was reduced so that it can be incorporated into an AAV vector for studies described later. This was carried out by reducing the size of the BPV^a region from 926 to 548 bp. The deletion reduces the total length of BPV sequences (BPV^a+BPV^b) from 2098 to 1720 bp. Rubnitz and Subramani (1984) have shown that the length of homology

between repeated sequences in a transfected plasmid is proportional to the recombination frequency between these sequences, so long as both sequences are over 200 bp in length. The above deletion is expected to cause a small decrease in the recombination frequency between the endogenous PV genome and 3S³5-dF as compared to the full-size 3S³5 construct. Fig. 12 shows that the smaller 3S³5-dF expression cassette does not express TK when cotransfected along with pUC19 DNA into COS-1 cells. The positive control plasmid S2.1TP-F containing a FLAG–TK fusion expressed high levels of TK (Fig. 12). To determine if 3S³5-dF can express TK in the presence of BPV DNA, COS-1 cells were cotransfected with 3S³5-dF and a BPV-1-containing plasmid. Western blot analysis indicated that cotransfection of 3S³5-dF with either circular or linear BPV-1 DNA resulted in TK expression (Fig. 12). There was a slight increase in TK expression in the presence of linear BPV-1 target, consistent with the possibility that homologous recombination of 3S³5-dF with linear BPV-1 DNA is more efficient.

The use of AAV to deliver a TK expression cassette to papillomavirus-infected cells

For a therapeutic expression cassette to be useful *in vivo*, an effective vector must be used for delivery. We, therefore, tested whether AAV, which is capable of infecting mucosal cells (such as those harboring PVs) can be used for the conditional expression of the TK gene. We created a modified version of the 3S³5-dF expression cassette, 3S5, in which two of the SV40 poly(A) sites were deleted from the SV40 poly(A) trimer to further reduce the size of the cassette. Our studies had suggested that splice site

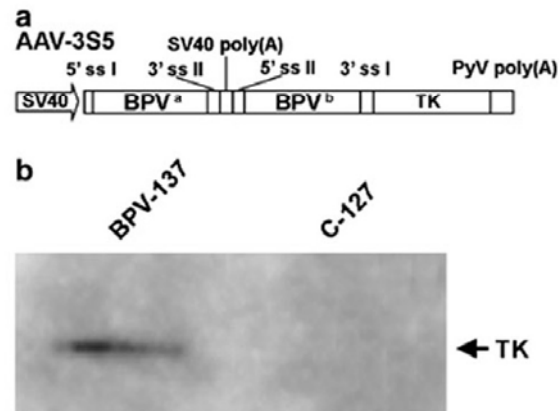
Fig 12. Recombination-dependent expression of FLAG-TK in cells transfected with various constructs



Protein lysates made from COS-1 cells transfected with the indicated plasmids were subjected to Western blot analysis using anti-FLAG M2 monoclonal antibodies. S2.1TP-F is a positive control expressing a FLAG-TK fusion. The 3S35-dF cassette was cotransfected into COS-1 cells along with either the negative control plasmid pUC19, the circular BPV plasmid, or a linearized BPV-1 plasmid (BPV-lin). The FLAG-TK fusion protein is indicated by an arrow. A GFP plasmid was transfected into COS-1 cells as a negative control.

placement, rather than the relative strength of the poly(A) site, largely determines the polyadenylation pattern in these expression cassettes (data not shown). Therefore, we postulated that there would be little or no effect of reducing the SV40 trimer to a monomer on TK expression. The 3S5 plasmid was found not to express TK as determined by Western blot analysis when introduced into PV-negative COS-1 cells by transfection (data not shown). We then incorporated the 3S5 cassette into the AAV vector, generating AAV-3S5 (Fig. 13a). The AAV-3S5 vector was used to infect BPV-137 cells, a C127 mouse fibroblast cell line that harbors about 1000 BPV-1 genomes per cell (McBride & Howley, 1991) or the parental C127 cells at a multiplicity of infection of 10 000 vector genomes per cell. As the vast majority of viral particles are not infectious in mouse cells (unpublished data), this corresponds to a transducing units of approximately 10 per cell. Since human cell lines containing stably replicating BPV-1 DNA are not available, we utilized the murine BPV-137 cell line in these experiments. This cell line was chosen because it harbors a BPV-1 mutant at a high copy number, which might increase the frequency of recombination over a cell line containing wild-type BPV-1. Since murine cell lines are poorly permissive for AAV infection, cells were pretreated with hydroxyurea to increase the transduction efficiency (Hansen *et al*, 2001). Hydroxyurea treatment at these low levels is thought to alter intracellular trafficking and enhance conversion of single-stranded AAV DNA to double-stranded genomes (Hansen *et al*, 2001). Also, cell cycle arrest is not seen at these low levels of HU (data not shown). Following infection, TK expression was assayed by Western blot analysis. To reduce the background signals, cell lysates were first immunoprecipitated with an anti-Flag M2-agarose affinity gel, and then subjected to

Fig 13. Recombination-dependant TK expression in BPV-positive cells upon infection with an AAV-TK vector



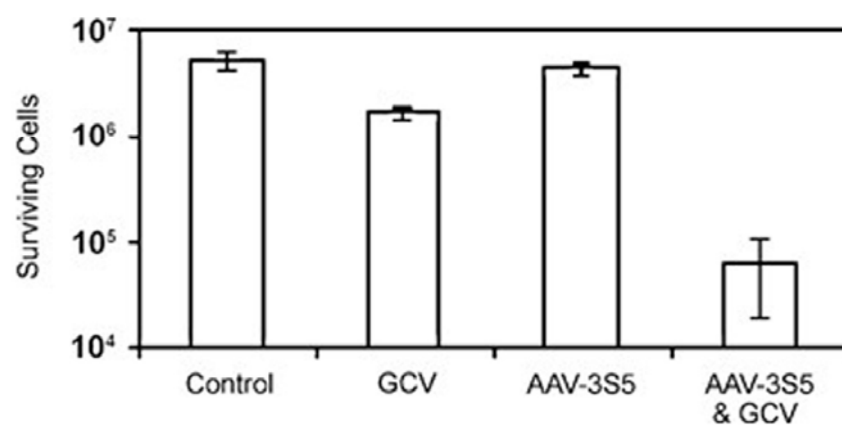
(a) Diagram of the AAV-3S5 construct used for TK expression. (b) Western blot analysis of TK expression upon AAV-3S5 infection. BPV-positive (BPV-137) or BPV-negative (C-127) murine cell lines were infected with the AAV-3S5 vector. Protein lysates were harvested 3 days postinfection, immunoprecipitated with anti-FLAG M2 monoclonal antibodies, and then precipitated proteins subjected to Western blot analysis using anti-TK polyclonal antibodies. The position of the TK band is indicated by an arrow.

Western blot analysis using anti-TK polyclonal antibodies. As shown in Fig. 13b, BPV-137 cells infected with AAV-3S5 expressed TK, while the parental BPV-negative cell line C-127 did not. This result is similar to that obtained in the cotransfection assay using TK cassettes along with a BPV-1 plasmid (Fig. 12). These results demonstrate that TK expression can be induced by DNA recombination in BPV-positive cells upon delivery of a nonexpressing cassette by an AAV vector.

The use of rAAV and ganciclovir to eliminate PV-infected cells

Although TK expression was detectable through the use of immunoprecipitation followed by Western blot analysis, we wished to determine whether TK levels were sufficient to eliminate PV-infected cells upon treatment with GCV. For this, BPV-137 cells were either mock infected or infected with AAV-3S5. The cells were then treated with GCV for a period of 4 days. Cells were harvested using trypsin and attached cells were counted. As shown in Fig. 14, both infected and uninfected BPV-137 cells rapidly proliferated without GCV treatment. BPV-137 cells that were uninfected but treated with GCV also proliferated rapidly and to nearly the same levels. On the other hand, GCV treatment resulted in over 95% cell death of AAV-3S5-infected cells (Fig. 14). Moreover, less than 65% of the remaining AAV-3S5/GCV-treated cells were viable as compared to over 95% viability of untreated control cells as assayed by trypan blue staining (data not shown). These results demonstrate that TK is expressed to levels high enough to sensitize PV harboring cells to killing by GCV treatment.

Fig 14. GCV inhibits the growth of BPV-positive cells infected with an AAV-TK vector



BPV-137 cells were either mock infected or infected with AAV-3S5 and cells were treated with 50 M GCV. Cells were harvested 4 days postinfection and counted.

3.4 DISCUSSION

We have developed a DNA recombination-based approach to eliminate PV infection. The efficacy of this approach is dependent on the ability of recombination to promote the expression of a suicide TK gene specifically in cells containing PV sequences. Using a variety of TK constructs, we demonstrate that TK expression can be specifically induced in PV-positive cells presumably as a result of homologous recombination between the cloned PV sequences with those present in the transfected/infected cells. This approach is expected to result in TK expression only in those cells in which a double crossover recombination event has occurred, resulting in reconstitution of the expression cassette. Single recombination events will not generate a functional transcription unit, and therefore are not expected to result in TK expression.

The expression of many natural genes is known to be regulated by alternative poly(A) site selection and depends upon both the position and relative strength of the poly(A) sites (reviewed in Edwards-Gilbert *et al*, 1997). Among identical poly(A) signals, the 5' site is favored over a more distal downstream site (Demone & Cole, 1988). Also, when a strong poly(A) site precedes a weak one by a short distance, less than 5% of the processed mRNAs terminate at the weak site (Batt *et al*, 1994). It was somewhat surprising; therefore, that TK expression in SS²TP was comparable to S2.1TP with the inclusion of strong upstream polyadenylation sites (Fig. 10). Cooke *et al* (1999) have shown that a 5' splice site, if not followed by a 3' splice site, can interfere with downstream polyadenylation. This interference may explain the high expression of TK by SS²TP. To eliminate this possibility, a 3' splice site was introduced upstream of the SV40 polyadenylation sites to create 3S³. The low levels of expression seen in 3S³ (Fig. 10)

may be due to splicing between 5' ss I and 3' ss I in a mechanism similar to exon skipping. Exon skipping has been shown to decrease when the 5' splice site following the skipped exon is strengthened, but increases when the 5' splice site is weakened or eliminated (Kuo *et al*, 1991). 3S³ contains no 5' splice site following the SV40 poly(A) signals, so one was introduced to create 3S³5, which does not express TK (Fig. 10).

We have utilized alternative poly(A) site selection as a means to regulate the expression of a suicide gene, TK, in cell lines. TK expression cassettes were developed in which strong SV40 poly(A) site(s) are surrounded by BPV-1 sequences (Fig. 9). Furthermore, the BPV-1 sequences are surrounded by 5' and 3' splice sites, respectively. Transcription initiated from the upstream SV40 early promoter is expected to generate a primary transcript, which will be spliced utilizing 5' ss I and 3' ss II, resulting in the termination of transcription and polyadenylation at the SV40 poly(A) site. Since the HSV-1 TK gene is located downstream of this region (Fig. 9), little or no expression of the TK gene is expected. This was indeed the case when the COS-1 cell line was transfected with the 3S³5 vector (Fig. 10 & 11). These results suggest that proper placement of the poly(A) sites as well as 5' and 3' splice sites in 3S³5 promotes efficient transcription termination, and therefore no detectable TK expression. We included two or three copies of the SV40 poly(A) sequence in some constructs since multiple copies of this sequence were previously found to block spurious transcription in a plasmid containing a promoter-less reporter gene (Maxwell *et al*, 1989).

We tested whether homologous recombination can be used as a means to promote TK expression exclusively in PV-positive cells. For this, we cloned a 2.1-kb region of the BPV-1 L1 and L2 ORFs (nt positions 4501–6647) into the various constructs.

Furthermore, the SV40 poly(A) sequence was inserted in the middle of the above region such that the upstream BPV^a region consisted of 926 bp, while the downstream BPV^b region contained 1169 bp (Fig. 9). The above region of the BPV-1 genome contains no known promoter element, and contains only part of the L1 and L2 ORFs and so it is not expected to express any protein. We hypothesized that when the above constructs are present in BPV-positive cells, homologous recombination between the L1/L2 regions present in the constructs and the BPV-1 genome in cells would result in deletion of the SV40 poly(A) sequence and regeneration of a TK expression cassette in which the polyomavirus poly(A) site present downstream of the TK gene is used for the termination of transcription (Fig. 9). Western blot analysis of COS-1 cells cotransfected with the 3S³5 and a BPV-1-containing plasmid showed expression of TK (Fig. 12). Since no TK was expressed in the absence of the BPV-1 plasmid (Fig 8), these results suggest that the most likely mechanism for the induction of TK expression involves homologous recombination between the L1/L2 regions of the two plasmids. This possibility is further supported by the slight increase in TK expression when the recombination target BPV-1 plasmid was linearized prior to transfection (Fig. 12). Homologous recombination is known to increase when one of the recombination substrates has linear ends (Ayers *et al*, 1985).

For a therapeutic expression cassette to be useful *in vivo*, an effective vector must be used for delivery. AAV is an ideal viral vector for this application since it commonly infects humans, but is not known to cause any human disease (Xiao *et al*, 1998). It has the ability to infect efficiently many types of dividing and nondividing cells (Xiao *et al*, 1998). Wild-type AAV-2 has also been found in significant numbers of cervical

epithelial cells from patients, indicating that AAV can infect cells harboring mucosa-tropic HPVs (Venturoli *et al*, 2001). Our experiments using the AAV-3S5 construct showed that TK is expressed in BPV-positive cells but not in BPV-negative cells (Fig. 13b). These results suggest that homologous recombination between the L1/L2 sequences in the TK cassette and the BPV-1 genome results in the expression of the TK gene specifically in BPV-1-positive cells.

The results of our studies demonstrate that homologous recombination can be used as a means to express a suicide gene, TK, in virally infected cells. Such cells are then sensitive to treatment with the well-established anti-herpesvirus drug GCV allowing elimination of over 95% of PV-infected cells (Fig. 14). This will allow us to develop a strategy in the future to eliminate specifically HPV-infected cells using GCV. Also, due to the well-known 'bystander effect' of TK, HPV-positive cells within a lesion that are not infected by the TK cassette are also likely to be eliminated by GCV treatment. An approach to eliminate PV-infected cells has been described by Sethi and Palefsky (2003), whereby TK expression is transcriptionally regulated by the presence of the viral E2 protein. We have developed an entirely different approach in which only the presence of viral DNA is required for TK expression.

Many high-risk HPVs such as types 16 and 18 are associated with the development of cancer of the cervix and oral cavity (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrmann & Laimins, 2003; zur Hausen, 2002). Frequently, the HPV DNA is integrated into the genome in cancers which express the viral E6 and E7 oncogenes to high levels (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrmann & Laimins, 2003; zur Hausen, 2002). The mechanisms and/or events that promote

integration of the HPV genome are not well known. Therefore, any approach for the elimination of high-risk HPVs from infected cells will have to be carefully developed so as not to promote integration of the viral DNA into the genome. However, low-risk HPVs such as types 6 and 11 cause long-term infections and painful lesions but are rarely associated with the development of cancers (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Their inability to promote malignant transformation is primarily due to the fact that their E6 and E7 oncoproteins do not inactivate the functions of the cellular tumor suppressor proteins such as p53 and pRB (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Thus, the strategy described in this report can be readily adapted for the treatment of lesions caused by low-risk HPVs. Also, since TK expression is limited to infected cells, this approach is likely to have little effect on uninfected cells. Finally, our recombination approach may also be potentially useful for the elimination of other viral infections.

Chapter 4

GENERAL DISCUSSION

4.1 CHARACTERIZATION OF TRANSCRIPTION FACTORS BINDING TO HPV-16 DURING CELLULAR DIFFERENTIATION

We have used the W12 clone 20863 to identify changes in TF binding to HPV during host cell differentiation. This cell line harbors HPV-16 episomally (Fig. 5) and can be induced to differentiate using standard techniques (Fig. 6). We used the Panomics transcription factor array to characterize the many changes in transcription factor binding during differentiation in this cell line. This analysis shows that the levels/DNA binding activity of more than 150 TFs changes during differentiation, out of approximately 345 TFs (Fig. 7). We also modified this technique to identify which of these transcription factors bind HPV-16, as well as to identify factors that bind known promoter regions. With this methodology, we have identified over 30 factors which may be involved in HPV-16 transcription (Fig. 7).

Many of these factors, such as C/EBP alpha, NF-1 and YY1, have been shown to interact with HPV-16 previously (Ai *et al*, 1999; Chan *et al*, 1990; Ciccolini *et al*, 1994; Gloss & Bernard, 1990). These known factors represent positive controls and partially validate the unknown factors identified by this analysis. Many other factors, such as NFkB, and Stat5, have been loosely associated with HPV infection or HPV related malignancies, but were not known to be directly involved in transcription of the viral genome (Bodily & Meyers, 2005; Hirokawa *et al*, 2003). These data may represent a mechanism by which these associated factors may affect papillomavirus expression during maturation or carcinogenesis. The majority of identified factors, such as C-Myb, Pax5, NF-ATx and WT1, represent novel interactions between transcription factors and the HPV-16 genome. Many transcription factors identified by this analysis, such as

C/EBP alpha, C-Myb, NF-ATx, and Pax5 have known roles in development and/or differentiation (Gloss & Bernard, 1990; Hadaschik *et al*, 2003; O'Conner *et al*, 2000; Rocha-Zavaleta *et al*, 2004). We expected changes in binding of many TF's involved in differentiation to change during the induction of the process of terminal differentiation in epithelial cells. Some identified transcription factors, such as NFkB and WT1, have been associated with cancers (Grassmann *et al*, 1996; Hebner & Laimins, 2005). A few identified transcription factors, such as NF-1, are involved in the transcription of a wide variety of promoters in many cell types under many conditions. While others, such as STAT5, activate promoters under more specific conditions (Hadaschik *et al*, 2003).

In order to validate data obtained from these arrays, a few genes of each class were further analyzed using consensus binding sequences obtained from the TRANSFAC database (Wingender *et al*, 2000). These binding sites were compared to the *Pst-I* fragment of HPV-16 which contains all putative promoters. Not all TF's identified by the Panomics array have identifiable binding sites in this region (data not shown). The majority of transcription factors have binding sites that map to known or putative promoter regions (Fig. 8). Of these factors, most binding sites map to regions associated with late gene expression. These poorly characterized promoters are thought to be dramatically upregulated during host cell differentiation as compared to the modest upregulation observed in the p97 promoter (Fehrmann & Laimins, 2003; Nees *et al*, 2001).

We have further validated binding as well as quantified increases in binding for these transcription factors using ChIP assays. These ChIP assays demonstrate that all TF's investigated bind HPV-16, and that binding increases in all factors upon

differentiation. In the transcription factors C/EBP alpha, C-Myb and NF1, binding increased by 8 fold or more. NF-ATx and Pax5 showed modest increases of about 2.6 and 1.9 fold, respectively. STAT5 and WT1 showed small but reproducible increases in binding of about 1.5 fold. This represents the first *in vivo* evidence that some of these factors bind HPV.

We have also used Western blot analysis to determine if increases in TF binding correspond to further validate increases in various TF at the protein level during differentiation. We see large increases in C-Myb and NF-1 protein levels during differentiation, which corresponds to large increases in viral binding as determined by ChIP assay and oligonucleotide binding as determined by Panomics array. The C/EBP alpha protein was substantially increased during differentiation, which matched well with the 7.6 fold increase in oligonucleotide binding and the 8.0 fold increase in viral binding. The STAT5 protein was also consistent across all analyses with a modest increase in the level of active phosphorylated STAT5 during differentiation, a 3.8 fold increase in oligonucleotide binding, and a small, but reproducible increase of 1.5 fold in viral binding. Levels of Pax5 and WT1 protein are also increased dramatically during differentiation. These data match well with the large increases seen in oligonucleotide binding as determined by Panomics array for Pax5 and WT1, but do not correlate strongly with the small increases in viral binding seen in the ChIP assay. This could be a result of the saturation of one or more TF binding sites during differentiation. For example, if half of the WT1 binding sites are occupied prior to differentiation, ChIP assays can detect a maximum of two-fold increase. The two-fold increase reported by the ChIP assay may differ substantially from the drastic increases in expression or activity

reported by Western blot or Panomics array, but it reveals the same trend. The NF-ATx protein was increased only slightly during differentiation as assayed by western blot, which contrasts with the dramatic increases by Panomics array and ChIP assay. This could be due to differences in the nuclear localization of NF-ATx, as NF-ATx activity is known to be regulated in this manner (Rocha-Zavaleta *et al*, 2004).

4.2 FUTURE DIRECTIONS IN THE STUDY OF HPV-16 TRANSCRIPTION

In the future, we plan to investigate the functional consequences of the TF/HPV interactions outlined in Chapter 3. We are currently in the process of creating reporter plasmids containing either the p97 or p670 promoters (see introduction) upstream of the luciferase reporter gene. Since some TF binding sites upstream of bp 97 might effect transcription of the p670 promoter, we are also creating a reporter cassette including both the p97 and p670 promoters in which luciferase is inserted downstream of the p670 promoter. As we are primarily interested in the activity of the late promoter in this construct, we will mutate the upstream TATA box to prevent initiation at the p97 promoter. This technique was first used in HPV-31 to create a reporter representing the late promoter with all upstream transcription factor binding sites (Spink & Laimins, 2005). These reporter plasmids will be transfected into HaCat cells (an HPV negative keratinocyte cell line), protein will be extracted before and after differentiation, and luciferase assays will be performed. As reported previously, transcription from the p97 promoter is expected to increase about 1.5 to 3 fold upon differentiation of the host cell (Apt *et al*, 1996; Grassmann *et al*, 1996; Yukawa *et al*, 1996). The activity of the p670 promoter, both short and full length, are expected to increase about 5-12 fold upon cellular differentiation (Grassmann *et al*, 1996; Rosensteirne *et al*, 2003).

Once these basal levels of induction are established, a series of mutations will abolish TF binding sites throughout the full length p670 promoter. We are interested in deleting the NF-1 sites in the full length p670 promoter. This is expected to reduce the basal levels of transcription before and after differentiation (Ai *et al*, 1999). The mutant

reporter may also exhibit a reduction in fold increase upon differentiation, which would be a strong indication that NF-1 is intimately involved in the regulation of viral gene expression. We have planned similar experiments with reporter plasmids in which C/EBP α , STAT5, C-Myb, and WT-1 sites are abolished. Following transfection, any reduction in the fold induction upon differentiation is a strong indication of the involvement of that particular factor in control of the life cycle of HPV-16. Additionally, we are planning to cotransfect the p97 and p670 promoters along with cassettes expressing various TFs. We expect to see an increase in luciferase expression if a particular TF stimulates transcription from these promoters. The overexpression of some factors may not result in an increase in p97/p670 expression, due to a variety of factors, but any increases in transcription are indications of the importance of TF binding. In order to further confirm binding of some of these factors, we also wish to overexpress and purify various TF proteins and test their binding to putative binding sites in wt promoters by EMSA. This type of confirmatory experiment will be especially important for factors with no previously known HPV involvement, such as C-Myb, STAT5, Pax5, and WT-1.

Recently, it has been shown that HPV DNA is methylated *in vivo* (Kalantari *et al*, 2004). We are interested in identifying functional consequences of this methylation and elucidating the role of methylation in the HPV life cycle. All seven NF-1 binding sites contain GC pairs which may be methylated and the Pax5 binding site contains four GC pairs. In order to determine if these sites may play a role in transcriptional regulation, we are planning to use bisulfite sequencing to determine if these sites are methylated *in vivo* and if this methylation pattern changes during differentiation. If a particular GC pair is

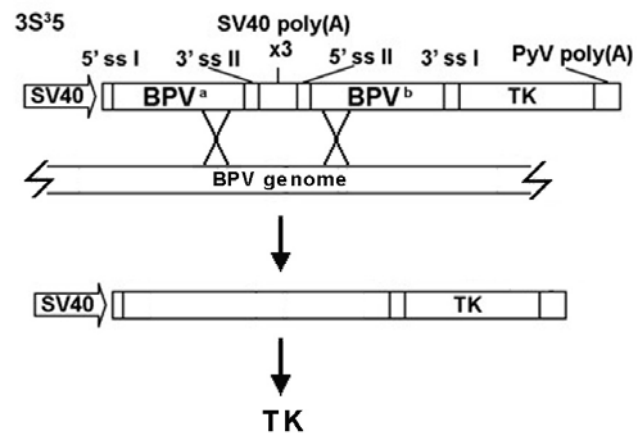
found to be differentially methylated, we plan to perform competitive EMSAs with both methylated and unmethylated substrates to determine if methylation affects the affinity of the TF towards DNA. We also plan to mutate the particular GC pairs in the previously mentioned p97 and p670 luciferase reporter cassettes and use transfection to determine the effect of these mutations on transcription.

4.3 OUTLOOK ON A RECOMBINATION-BASED APPROACH TO ELIMINATE PAPILLOMAVIRUS INFECTION

Stratagem

We have developed a DNA recombination-based approach to eliminate PV infection. This system enables the use of an untargeted viral vector, AAV, to deliver the cassette *in vivo*. Using a variety of TK constructs, we have demonstrated that TK expression can be specifically induced in PV-positive cells. The nature of the expression cassette limits TK expression to cells in which homologous recombination occurs between segments of the viral genome and the expression cassette. As shown in Fig. 15, a recombinatorial event resulting in removal of the premature polyadenylation site and flanking splice sites results in alternative polyadenylation site selection and alternate splicing and leads to TK expression. In order to have specificity of the induction of TK, only the result of a double cross-over homologous recombination event or a gene-conversion event is designed to express TK. These and other likely products of DSB repair are evaluated case by case below for 3S³5-based expression cassettes to illustrate the basis and rational of this strategy.

Fig 15. Model of homologous recombination leading to TK expression



Case 1: No recombination

If there is no recombination, there will be no TK expression. Transcription is initiated at the SV40 promoter. Splicing occurs between 5' splice site I and 3' splice site II to remove the BPV^a region, and polyadenylation occurs at one of the SV40 poly(A) signals. As mentioned in the introduction, the SV40 poly(A) sites are favored over the downstream polyomavirus poly(A) site as a result of the strength of the sites, the distance to the polyomavirus site, and the placement of the splice sites. The TK ORF will not be included in the resulting message, and no TK will be expressed. This is confirmed by Northern and Western blot analysis shown in Fig. 10 & 11.

Case 2: Double cross-over homologous recombination between BPV-derived regions and the BPV genome

If the expression cassette and the BPV genome are both present in the cell, a double cross-over homologous recombination event or other gene conversion event can occur between homologous regions in the introduced expression cassette and the BPV genome. This can trigger TK expression. In the recombined construct, the SV40 poly(A) signal will be removed from the construct so that the BPV regions are continuous. Transcription in this cassette (modeled as S2.1TP) starts at the SV40 promoter. Splicing occurs between 5' splice site I and 3' splice site II, removing BPV derived regions from the resulting transcript. Transcription also proceeds through the TK ORF, and the polyomavirus polyadenylation signal is utilized. This pattern of expression is confirmed by Northern and Western blot analysis shown in Fig. 10 & 11.

Case 3: Single cross-over homologous recombination events between the expression cassette and BPV genome

Constructs that recombine at a single site will not express TK since the SV40 promoter will be separated from the TK ORF. It would be attractive to trigger TK expression through single cross-over, since these events will be specific to PV infected cells and will have a higher frequency than the double cross-over events. However, it is difficult to design strategies that trigger TK expression from a cassette as a result of a single cross-over event. Recombination events of this type create fusions of the introduced expression cassette and the PV genome with PV sequences either upstream or downstream (depending on the design) of the TK ORF. One way in which this could trigger TK expression is that downstream PV sequences might be chosen that contain a poly(A) signal which could replace plasmid sequences containing a strong poly(A) signal. Because weak poly(A) signals still result in substantial expression, this change is not likely to tightly regulate TK expression. Alternatively, a single cross-over event might be designed to replace plasmid sequences upstream of the TK ORF with PV sequences. If the plasmid sequence undergoing recombination did not have a strong promoter activity (and the PV sequence contained a strong promoter) the recombined expression cassette would express more TK than the expression cassette prior to recombination. Unfortunately, the strongest PV promoter is much weaker than the strong CMV or SV40 promoters used in mammalian expression cassettes. Also, an expression cassette designed to trigger TK expression when under the control of a strong promoter might not be highly specific to PV infected cells. A cassette of this type could integrate

through non-homologous recombination into the host cell genome behind a strong promoter. These types of events are fairly common in many cell types, and would lead to TK expression from PV-negative cells, resulting in a loss of specificity.

Case 4: Nonhomologous recombination involving the expression cassette and the host cell genome

Since nonhomologous recombination events do not require extensive homology, any sequence from the expression cassette that picks up a DSB could be repaired by ligation (by either single-stranded annealing or nonhomologous end-joining mechanisms) to virtually any genomic sequence. This could yield either partial or whole inclusions of the expression cassette in the genome of the cell. Partial inclusions of the transcriptional unit would likely not express TK, as they would separate the SV40 promoter from the TK ORF. Likewise, whole inclusions of the transcriptional unit will not express TK, as the premature SV40 poly(A) site will be preferred.

It should be noted that a nonhomologous recombination event resulting in the insertion of the TK ORF (lacking promoter/regulatory sequences) downstream of an active promoter could result in TK expression in PV-negative cells. As PV-negative cells infected or transfected with the non-expressing cassette do not express detectable levels of TK, we can conclude that this type of event is very rare. Consequently, although this type of event remains a theoretical possibility, we are still confident that very, very few PV-negative cells will express TK by this mechanism.

Random integration of rAAV in various places throughout the host cell genome does have the potential to disrupt the expression of tumor suppressors and has some

limited oncogenic potential. This is a very real safety concern. It should be noted that there is no record of rAAV or wt AAV causing cancer, even though most people are seropositive for at least one type of AAV. HPV, on the other hand, has a very strong and well documented association with cancer. In summary, the risk of using a virus like rAAV to eliminate PV infection, while not zero, is lower than the risk of continued high risk HPV infection.

Case 5: Nonhomologous recombination events within the expression cassette

Events of this type will create deletions and junctional aberrations in the expression vector. Most deletions will not result in TK expression, such as a deletion in the SV40 promoter, 5' splice site I, or the TK ORF, etc. A deletion in the SV40 poly(A) signals without deletions in 5' splice site I or 3' splice site II will result in TK expression for the same reasons listed in Case 2. This type of event would occur in cells infected/transfected with TK expression cassette in PV-negative cells, Fig. 10 & 11 confirm that this event is undetectable.

Design of expression cassettes

We have utilized alternative poly(A) site-selection as a means to regulate the expression of a suicide gene, TK, from a gene cassette in cell lines. The desirable features of this gene cassette are that it must not express TK, but in the presence of PV sequences in the cell, it should undergo homologous recombination leading altered poly(A) signal usage and TK expression. The expression of many natural genes is known to be regulated by alternative poly(A) site selection and this selection is known to be

dependent on both the position and relative strength of the poly(A) sites (reviewed in section 1.5). Among identical poly(A) signals, the 5' site is favored over a more distal downstream site (Batt *et al*, 1994). Also, stronger poly(A) sites can be used with much higher frequency, whether they are upstream or a short distance downstream of the weaker site.

We generated the SS²TP expression cassette (Fig. 9) to take advantage of these known criteria for poly(A) site selection. In this expression cassette, an intron containing segments of the BPV genome was inserted upstream of the TK ORF and the weak late poly(A) site of polyomavirus (Batt *et al*, 1994). Two SV40 poly(A) sites were then inserted into the middle of the above BPV sequence (Fig. 9). We anticipated that transcripts from this expression cassette would primarily terminate at the strong upstream SV40 poly(A) sites. It was somewhat surprising; therefore, that TK expression in SS²TP was comparable to S2.1TP even with the inclusion of strong upstream polyadenylation sites (Fig. 10 & 11).

Cooke *et al* (1999) have shown that a 5' splice site, if not followed by a 3' splice site, can interfere with downstream polyadenylation. This interference may explain the high expression of TK by SS²TP. To eliminate this possibility, a 3' splice site was introduced upstream of the SV40 polyadenylation sites to create 3S³. The low levels of expression seen in 3S³ (Fig. 10) may be due to splicing between 5' ss I and 3' ss I in a mechanism similar to exon skipping. Exon skipping has been shown to decrease when the 5' splice site following the skipped exon is strengthened, but increases when the 5' splice site is weakened or eliminated (Kuo *et al*, 1991). 3S³ contains no 5' splice site following the SV40 poly(A) signals, so one was introduced to create 3S³5, which does

not express TK (Fig. 10). These results suggest that proper placement of the poly(A) sites as well as 5' and 3' splice sites in 3S³5 promotes efficient transcription termination, and therefore no detectable TK expression.

Ability to regulate TK expression upon cotransfection

We tested whether homologous recombination can be used as a means to promote TK expression exclusively in PV-positive cells. For this, we cloned a 2.1-kb region of the BPV-1 into the various constructs for use as a recombination target and cotransfected this construct along with the 3S³5 plasmid. We hypothesized that when the above constructs are present in BPV-positive cells, homologous recombination could occur between the BPV derived regions present in the constructs. This recombination could result in deletion of the SV40 poly(A) sequence and result in a TK expression cassette in which the downstream polyomavirus poly(A) site is used, and TK is expressed. Western blot analysis of COS-1 cells cotransfected with 3S³5 and a BPV-1-containing plasmid showed expression of TK (Fig. 12). Since no TK was expressed in the absence of the BPV derived sequences (Fig. 12), these results suggest that TK expression involves homologous recombination between the L1/L2 regions of the two plasmids as designed. This possibility is further supported by the slight increase in TK expression when the recombination target BPV-1 plasmid was linearized prior to transfection (Fig. 12). Homologous recombination is known to increase when one of the recombination substrates has linear ends (Ayers *et al*, 1985).

Effective inhibition of replication in tissue culture cells harboring BPV is fundamentally a function of the number of cells in which a productive recombinatorial

event takes place and subsequently express TK. This number of cells is of particular concern because it represents the minimum number of infected cells in which replication will be inhibited by ganciclovir treatment. The actual number of inhibited cells is likely to be somewhat higher because phosphorylated versions of ganciclovir can leak through gap junctions from cells that express TK to prevent replication in neighboring cells (Dilber *et al*, 1997). It is, however, potentially interesting to determine the percentage of cells that express TK in our cotransfection system (Fig. 12), as well as the percentage of plasmids which have undergone recombination in the same system. Work attempting to quantify the percentage of recombined expression cassettes in this cotransfection system has not been highly informative thus far. These assays were performed by first isolating low-molecular weight DNA from co-transfected and control cells through Hirt fractionation (Hirt, 1967). Southern Blots were performed directly on these Hirt fractions, but the results were inconclusive, as recombinant digestion patterns were obscured by the vast excess of unrecombined plasmids (data not shown). Because the recombined expression cassette acts as plasmid, these Hirt fractions were then used to transform *E. coli*, and plasmid DNA from the mixed population of transformed bacteria was harvested. This plasmid DNA was then digested with a restriction enzyme designed to digest the unrecombined plasmid, but leave the recombined plasmid intact. Fresh *E. coli* was transformed with this digested DNA, individual clones were isolated, and their plasmid DNA was screened for similarity to the predicted recombination product. This process has yielded plasmids which match recombined plasmid, according to restriction enzyme analysis. This process cannot, however, determine the frequency of recombination and was not pursued in depth.

Ability to Regulate TK Expression when Delivered by AAV

For a therapeutic expression cassette to be useful *in vivo*, an effective vector must be used for delivery. AAV is an ideal viral vector for this application since it commonly infects humans, but is not known to cause any human disease (Hirata & Russel, 2000). It has the ability to infect efficiently many types of dividing and nondividing cells (Hirata & Russel, 2000). Wild-type AAV-2 has also been found in significant numbers of cervical epithelial cells from patients, indicating that AAV can infect cells harboring mucosa-tropic HPVs (Venturoli *et al*, 2001). Our experiments using the AAV-3S5 construct showed that TK is expressed in BPV-positive cells but not in BPV-negative cells (Fig. 13b). These results suggest that homologous recombination between the L1/L2 sequences in the TK cassette and the BPV-1 genome results in the expression of the TK gene specifically in BPV-1-positive cells.

The results of our studies demonstrate that homologous recombination can be used as a means to express a suicide gene, TK, in virally infected cells. Such cells are then sensitive to treatment with the well-established anti-herpesvirus drug GCV, allowing elimination of over 95% of PV-infected cells (Fig. 14). This will allow us to develop a strategy in the future to eliminate specifically HPV-infected cells using GCV. Also, due to the well-known 'bystander effect' of TK, HPV-positive cells within a lesion that are not infected by the TK cassette are also likely to be eliminated by GCV treatment.

An approach to eliminate PV-infected cells has been described by Sethi and Palefsky (2003), whereby TK expression is transcriptionally regulated by the presence of the viral E2 protein. We have developed an entirely different approach in which only the presence of viral DNA is required for TK expression.

4.4 FUTURE OF RECOMBINATION-BASED APPROACHES TO ELIMINATE VIRAL INFECTIONS

Many high-risk HPVs such as types 16 and 18 are associated with the development of cancer of the cervix and oral cavity (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Frequently, the HPV DNA is integrated into the host cell genome in late stage cancers associated with these viruses (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Low-risk HPVs such as types 6 and 11 cause long-term infections and painful lesions, but are rarely associated with the development of cancers (Dell & Gaston, 2001; Stubenrauch & Laimins, 1999; Fehrman & Laimins, 2003; zur Hausen, 2002). Because the recombination based approach described is likely to not be capable of eliminating PV-positive cells in which the viral genome is integrated, this strategy would be best adapted for the treatment of lesions caused by low-risk HPVs or very early stage high-risk types. Also, since TK expression is limited to infected cells, this approach is likely to have little effect on uninfected cells. Finally, our recombination approach may also be potentially useful for the elimination of other viral infections.

In order to validate that the system can be adapted to other viruses, we plan to move the system from a BPV system into a HPV system. This involves the replacement of the BPV recombination regions (Fig. 9) with HPV sequences. These HPV sequences must not contain promoter elements, splice sites, or polyadenylation signals, as these may interfere with the regulatory mechanism. This construct would be tested in a cotransfection system (as with BPV) prior to the creation of a rAAV to deliver the construct. rAAV would also be used to deliver the therapeutic cassette into HPV positive cells and TK expression and activity would be assayed.

Chapter 5

BIBLIOGRAPHY

- Adams M, Wong C, Wang D & Romeo J.** 1999. Limitation of Tat-Associated transcriptional processivity in HIV-infected PBMC. *Virology*. **257**:397-405.
- Ai W, Toussaint E & Roman A.** 1999. CCAAT displacement protein binds to and negatively regulates human papillomavirus type 6 E6, E7, and E1 promoters. *J Virol*. **73**:4220-4229.
- Apt D, Chong T, Liu Y & Bernard HU.** 1993. Nuclear factor I and epithelial cell-specific transcription of human papillomavirus type 16. *J Virol*. **67**:4455-4463.
- Apt D, Watts RM, Suske G & Bernard HU.** 1996. High Sp1/Sp3 ratios in epithelial cells during epithelial differentiation and cellular transformation correlate with the activation of the HPV-16 promoter. *Virology*. **224**:281-291.
- Ayares D, Spencer J, Schwartz F, Morse B & Kucherlapati R.** 1985. Homologous recombination between autonomously replicating plasmids in mammalian cells. *Genetics*. **111**:375-388.
- Batt DB, Luo Y & Carmichael GG.** 1994. Polyadenylation and transcription termination in gene constructs containing multiple tandem polyadenylation signals. *Nucleic Acids Res*. **22**:2811-2816.
- Bodily JM & Meyers C.** 2005. Genetic analysis of the human papillomavirus type 31 differentiation-dependent late promoter. *J Virol*. **79**:3309-3321.
- Caruso M, Salomon B, Zhang S, Brisson E, Clavel F, Lowy I & Klatzmann D.** 1995. Expression of a Tat-inducible Herpes Simplex Virus thymidine kinase gene protects acyclovir-treated CD4 cells from HIV-1 spread by conditional suicide and inhibition of reverse transcription. *Virology*. **206**:495-503.
- Chan WK, Chong T, Bernard HU & Klock G.** 1990. Transcription of transforming genes of the oncogenic human papillomavirus type-16 is stimulated by tumor promoters through AP1 binding sites. *J Virol*. **18**:763-769.
- Chang YE & Laimins LA.** 2000. Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. *J Virol*. **74**:4174-4182.
- Chiang CM, Dong G, Broker TR & Chow LT.** 1992. Control of human papillomavirus type 11 origin of replication by the E2 family of transcription regulatory proteins. *J Virol*. **66**:5224-5231.
- Chong T, Apt D, Gloss B, Isa M & Bernard HU.** 1991. The enhancer of human papillomavirus type 16: Binding sites for the ubiquitous transcription factors Oct-1, NFA, TEF-2, NF1, and AP1 participate in epithelial cell-specific transcription. *J Virol*. **65**:5933-5943.

- Ciccolini F, Di Pasquale G, Carlotti F, Crawford L & Tommasino M.** 1994. Functional studies of E7 proteins from different HPV types. *Oncogene*. **9**:2633-2638.
- Cooke C, Hans H & Alwine JC.** 1999. Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol*. **19**:4971-4979.
- Dell G & Gaston K.** 2001. Contributions in the domain of cancer research: review of human papillomaviruses and their role in cervical cancer. *Cell Mol Life Sci*. **58**:1923-1942.
- Demone RM & Cole CN.** 1988. Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. *Mol Cell Biol*. 1988; **8**:4829-4839.
- Dilber MS, Abedi MR, Christensson B, Bjorkstrand B, Kidder GM, Naus CC, Gahrton G & Smith CI.** 1997. Gap junctions promote the bystander effect of herpes simplex virus thymidine kinase in vivo. *Cancer Res*. **57**:1523-1528.
- Dong XP, Stubenrauch F, Beyer-Finkler E & Pfister H.** 1994. Prevalence of deletions of YY1-binding sites in episomal HPV 16 DNA from cervical cancers. *Int J Cancer*. **58**:803-808.
- Edwards-Gilbert G, Veraldi KL & Milcarek C.** 1997. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res*. **25**:2547-2561.
- Fehrmann F & Laimins LA.** 2003. Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. *Oncogene*. **22**:5201-5207.
- Finzer P, Kuntzen C, Soto U, zur Hausen H & Rosl F.** 2001. Inhibitors of histone deacetylase arrest cell cycle and induce apoptosis in cervical carcinoma cells circumventing human papillomavirus oncogene expression. *Oncogene*. **20**:4767-4776.
- Flores ER & Lambert PF.** 1997. Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle. *J Virol*. **71**:7167-7179.
- Fukui T, Hayashi Y, Kagami H, Yamamoto N, Fukuhara H, Tohnai I, Ueda M, Mizuno M & Yoshida J.** 2001. Suicide gene therapy for human oral squamous cell carcinoma cell lines with adeno-associated virus vector. *Oral Oncol*. **37**:211-215.

- Gloss B & Bernard HU.** 1990. The E6/E7 promoter of human papillomavirus type 16 is activated in the absence of E2 proteins by a sequence-aberrant SP1 distal element. *J Virol.* **64**:5577-5584.
- Gopalakrishnan V & Khan SA.** 1994. E1 protein of human papillomavirus type 1a is sufficient for initiation of viral DNA replication. *PNAS.* **91**:9597-9601.
- Grassmann K, Rapp B, Maschek H, Petry KU & Iftner T.** 1996. Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *J Virol.* **70**:2339-2349.
- Gross-Mesilaty S, Reinstein E, Bercovich B, Tobias KE, Schwartz AL, Kahana C & Ciechanover A.** 1998. Basal and human papillomavirus E6 oncoprotein-induced degradation of Myc proteins by the ubiquitin pathway. *PNAS.* **95**:8058-8063.
- Haber JE.** 1999. DNA repair: Gatekeepers of recombination. *Nature Medicine.* **398**:665-667.
- Haber JE.** 2000. Partners and pathways: repairing a double-strand break. *Trends in Genetics.* **16**:259-264.
- Hadaschik D, Hinterkeuser K, Oldak M, Pfister HJ & Smola-Hess S.** 2003. The Papillomavirus E2 protein binds to and synergizes with C/EBP factors involved in keratinocyte differentiation. *J Virol.* **77**: 5253-5265.
- Hansen J, Qing K & Srivastava A.** 2001. Adeno-associated virus type 2-mediated gene transfer: altered endocytic processing enhances transduction efficiency in murine fibroblasts. *J Virol.* **75**:4080-4090.
- Havard L, Rahmouni S, Boniver J & Delvenne P.** 2005. High levels of p105 (NFKB1) and p100 (NFKB2) proteins in HPV16-transformed keratinocytes: role of E6 and E7 oncoproteins. *Virology.* **331**:357-366.
- Hebner CM & Laimins LA.** 2005. Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity. *Rev Med Virol.* **16**:83-97.
- Hirata HK & Russell DW.** 2000. Design and packaging of adeno-associated virus gene targeting vectors. *J Virology.* **74**:4612-4620.
- Hirokawa S, Sato H, Kato I & Kudo A.** 2003. EBF-regulating Pax5 transcription is enhanced by STAT5 in the early stage of B cells. *Eur J Immunol.* **7**:1824-1829.
- Hirt B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J Mol Biol.* **26**:365-369.

- Holt SE, Schuller G & Wilson VG.** 1994. DNA binding specificity of the bovine papillomavirus E1 protein is determined by sequences contained within an 18-base-pair inverted repeat element at the origin of replication. *J Virol.* **68**:1094-1102.
- Holt SE & Wilson VG.** 1995. Mutational analysis of the 18-base-pair inverted repeat element at the bovine papillomavirus origin of replication: identification of critical sequences for E1 binding and in vivo replication. *J Virol.* **69**:6525-6532.
- Hubert WG.** 2005. Variant upstream regulatory region sequences differentially regulate human papillomavirus type 16 DNA replication throughout the viral life cycle. *J Virol.* **79**:5914-5922.
- Kalantari M, Calleja-Macias IE, Tewari D, Hagmar B, Lie K, Barrera-Saldana HA, Wiley DJ & Bernard HU.** 2004. Conserved methylation patterns of human papillomavirus type 16 DNA in asymptomatic infection and cervical neoplasia. *J Virol.* **78**:12762-12772.
- Kitamura Y, Yoshikura H & Kobayashi I.** 1990. Homologous recombination in a mammalian plasmid. *Mol Gen Genet.* **222**:185-191.
- Kiyono T, Hiraiwa A, Fujita M, Hayashi Y, Akiyama T & Ishibashi M.** 1997. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. *PNAS.* **94**:11612-11616.
- Kuhne C & Banks L.** 1998. E3-ubiquitin ligase/E6-AP links multicopy maintenance protein 7 to the ubiquitination pathway by a novel motif, the L2G box. *J Biol Chem.* **273**:34302-34310.
- Kukimoto I, Aihara S, Yoshiike K & Kanda T.** 1998. Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. *Biochem Biophys Res Commun.* **249**:258-262.
- Kuo HC, Nasim FH & Grabowski PJ.** 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science.* **251**: 1045-1050.
- Lee SB & Haber DA.** 2001. Wilms tumor and the WT1 gene. *Exp Cell Res.* **264**: 74-99.
- Lee SS, Weiss RS & Javier RT.** 1997. Binding of human virus oncoproteins to hDlg/SAP97, a mammalian homolog of the Drosophila discs large tumor suppressor protein. *PNAS.* **94**:6670-6675.
- Maxwell IH, Harrison GS, Wood WM & Maxwell F.** 1989. A DNA cassette containing a trimerized SV40 polyadenylation signal which efficiently blocks spurious plasmid-initiated transcription. *BioTechniques.* **7**:276-280.

McBride AA & Howley PM. 1991. Bovine papillomavirus with a mutation in the E2 serine 301 phosphorylation site replicates at a high copy number. *J Virol.* **65**:6528-6534.

McIlwraith MJ, Van Dyck E, Masson JY, Stasiak AZ, Stasiak A & West SC. 2000. Reconstitution of the strand invasion step of double-strand break repair using human Rad51 Rad52 and RPA proteins. *J Mol Biol.* **304**:151-164.

Meyers C, Alam S, Mane M & Hermonat PL. 2000. Altered biology of adeno-associated virus type 2 and human papillomavirus during dual infection of natural host tissue. *Virology.* **287**:30-39.

Mizuno M, Yoshida J, Colosi P & Kurtzman G. 1998. Adeno-associated virus vector containing the herpes simplex virus thymidine kinase gene causes complete regression of intracerebrally implanted human gliomas in mice, in conjunction with ganciclovir administration. *Jpn J Cancer Res.* **89**:76-80.

Nakagawa S & Huibregtse JM. 2000. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. *Mol Cell Biol.* **20**:8244-8253.

Nees M, Geoghegan JM, Hyman T, Frank S, Miller L & Woodworth CD. 2001. Papillomavirus type 16 oncogenes downregulate expression of interferon-responsive genes and upregulate proliferation-associated and NF-kappaB-responsive genes in cervical keratinocytes. *J Virol.* **75**:4283-4296.

Nickoloff JA & Reynolds RJ. 1990. Transcription stimulates homologous recombination in mammalian cells. *Mol Cell Bio.* **10**:4837-4845.

O'Connor MJ, Stunkel W, Koh CH, Zimmermann H & Bernard HU. 2000. The differentiation-specific factor CDP/Cut represses transcription and replication of human papillomaviruses through a conserved silencing element. *J Virol.* **74**:401-410.

Ozbun MA. 2002. Human papillomavirus type 31b infection of human keratinocytes and the onset of early transcription. *J Virol.* **76**:11291-11300.

Ozbun MA & Meyers C. 1999. Two novel promoters in the upstream regulatory region of human papillomavirus type 31b are negatively regulated by epithelial differentiation. *J Virol.* **73**:3505-3510.

Pattison S, Skalnik DG & Roman A. 1997. CCAAT displacement protein, a regulator of differentiation-specific gene expression, binds a negative regulatory element within the 5' end of the human papillomavirus type 6 long control region. *J Virol.* **71**: 2013-2022.

- Patel D, Huang SM, Baglia LA & McCance DJ.** 1999. The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. *EMBO J.* **18**:5061-5072.
- Rocha-Zavaleta L, Huitron C, Caceres-Cortes JR, Alvarado-Moreno JA, Valle-Mendiola A, Soto-Cruz I, Weiss-Steider B & Rangel-Corona R.** 2004. Interleukin-2 (IL-2) receptor-beta gamma signalling is activated by c-Kit in the absence of IL-2, or by exogenous IL-2 via JAK3/STAT5 in human papillomavirus-associated cervical cancer. *Cell Signal.* **16**:1239-1247.
- Rosenstierne MW, Vinther J, Hansen CN, Prydsøe M & Norrild B.** 2003. Identification and characterization of a cluster of transcription start sites located in the E6 ORF of human papillomavirus type 16. *J Gen Virol.* **84**:2909-2920.
- Rouet P, Smith F & Jasin M.** 1994. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *PNAS.* **91**:6064-6068.
- Rubnitz J & Subramani S.** 1984. The minimum amount of homology required for homologous recombination in mammalian cells. *Mol Cell Biol.* **4**:2253-2258.
- Rushton JJ, Davis LM, Lei W, Mo X, Leutz A & Ness SA.** 2003. Distinct changes in gene expression induced by A-Myb, B-Myb and c-Myb proteins. *Oncogene.* **22**:308-313.
- Russell DW, Alexander IE & Miller AD.** 1995. DNA synthesis and topoisomerase inhibitors increase transduction by adeno-associated virus vectors. *PNAS.* **92**:5719-5723.
- Salomon B, Maury S, Loubiere L, Caruso M, Onclercq R & Klatzmann D.** 1995. A truncated Herpes Simplex Virus thymidine kinase phosphorylates thymidine and nucleoside analogs and does not cause sterility in mice. *Mol Cell Bio.* **15**:5322-5328.
- Sambrook J, Fritsch EF & Maniatis T.** 1989. *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press: Cold Spring Harbor.
- Scheffner M, Huibregtse JM, Vierstra RD & Howley PM.** 1993. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell.* **75**:495-505.
- Seidman MM.** 1987. Intermolecular homologous recombination between transfected sequences in mammalian cells is primarily nonconservative. *Mol Cell Bio.* **7**:3561-3565.
- Sen E, Alam S & Meyers C.** 2004. Genetic and biochemical analysis of cis regulatory elements within the keratinocyte enhancer region of the human

papillomavirus type 31 upstream regulatory region during different stages of the viral life cycle. *J Virol.* **78**:612-629.

Sen E, Bromberg-White JL & Meyers C. 2002. Genetic analysis of cis regulatory elements within the 5' region of the human papillomavirus type 31 upstream regulatory region during different stages of the viral life cycle. *J Virol.* **76**:4798-4809.

Serfling E, Berberich-Siebelt F, Chuvpilo S, Jankevics E, Klein-Hessling S, Twardzik T & Avots A. 2000. The role of NF-AT transcription factors in T cell activation and differentiation. *Biochim Biophys Acta.* **1498**:1-18.

Sethi N & Palefsky J. 2003. Treatment of human papillomavirus (HPV) type 16-infected cells using herpes simplex virus type I thymidine kinase-mediated gene therapy transcriptionally regulated by the HPV E2 protein. *Hum Gene Ther.* **14**:45-57.

Spink KM & Laimins LA. 2005. Induction of human papillomavirus type 31 late promoter requires differentiation but not DNA amplification. *J Virol.* **79**:4918-4926.

Stubenrauch F & Laimins LA. 1999. Human papillomaviruses: active and latent phases. *Cancer Biol.* **9**:379-386.

Sun LW, Li J & Xiao X. 2000. Overcoming the size limitation of adeno-associated viral vectors by heterodimerization. *Nature Medicine.* **6**:599-602.

Sun YN, Lu JZ-J & McCance DJ. 1996. Mapping of HPV-11 E1 binding site and determination of other important cis elements for replication of the origin. *Virology.* **216**:219-222.

Sverdrup F & Khan SA. 1994. Replication of human papillomavirus (HPV) DNAs supported by the HPV type 18 E1 and E2 proteins. *J Virol.* **68**:505-509.

Taghian DG & Nickoloff JA. 1997. Chromosomal double strand breaks induce gene conversion at high frequency in mammalian cells. *Mol Cell Bio.* **17**:6386-6393.

Thyagarajan B, McCormick-Graham M, Romero DP & Campbell C. 1996. Characterization of homologous DNA recombination activity in normal and immortal mammalian cells. *Nucleic Acids Res.* **24**:4084-4091.

Tsai TC & Chen SL. 2003. The biochemical and biological functions of human papillomavirus type 16 E5 protein. *Arch Virol.* **148**:1445-1453.

Van Horn G, Sheikh S & Khan SA. 2001. Regulation of Human papillomavirus type 1 replication by the viral E2 protein. *Virology.* 2001; **287**:214-224.

Venturoli S, Cricca M, Bonvicini F, Gallinella G, Gentilomi G, Zerbini M & Musiani M. 2001. Detection of adeno-associated virus DNA in female genital samples by PCR-ELISA. *J Med Virol.* **64**:577-582.

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I & Schacherer F. 2000. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* **28**:316-319.

Wilson VG, West M, Woytek K & Rangasamy D. 2002. Papillomavirus E1 proteins: form, function, and features. *Virus Genes.* **24**:275-290.

Yukawa K, Butz K, Yasui T, Kikutani H & Hoppe-Seyler F. 1996. Regulation of human papillomavirus transcription by the differentiation-dependent epithelial factor Epoc-1/skn-1a. *J Virol* **70**:10-16.

Xiao X, Li J & Samulski RJ. 1998. Production of high-titer adeno-associated virus vectors in the absence of helper adenovirus. *J Virol.* **72**:2224-2232.

Zimmermann H, Degenkolbe R, Bernard HU & O'Connor MJ. 1999. The human papillomavirus type 16 E6 oncoprotein can downregulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J Virol.* **73**:6209-6219.

Zolotukhin S, Byrne BJ, Mason E, Zolotukhin I, Potter M, Chesnut K, Summerford C, Samulski RJ & Muzyczka N. 1999. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Therapy.* **6**:973-985.

zur Hausen H. 2002. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer.* **2**:342-350.